5. Mentale Zustände

5.1. Mentale Zustände und Intentionalität

5.1.1. Begriffsklärung

Das Verhalten kognitiver Agenten (Menschen oder künstliche Agenten) basiert auf der Interaktion einer Anzahl von kognitiven Komponenten, die als *mentale Zustände* bezeichnet werden. Sie sind im „Kopf“ des Agenten vorhanden. Mentale Zustände sind Glauben (Annahmen), Wünsche, Intentionen und Versprechen. Mentale Zustände beziehen sich auf etwas Konkretes, sie haben einen *intentionalen* Charakter. Man glaubt etwas oder nimmt etwas an, man wünscht etwas oder man verspricht etwas. Sie können sich auch auf andere mentale Zustände beziehen, z.B. wenn man annimmt, dass Hans eine Tafel Schokolade wünscht, dann bezieht sich der eigene mentale Zustand des Annehmens auf den mentalen Zustand des Wünschens von Hans.

5.1.2. Kognitons

Da der Begriff *Zustand* im Zusammenhang mit mentalen Zuständen Schwierigkeiten macht, wird ein Kunstwort eingeführt, das *Kogniton*. Ein Kogniton bezeichnet eine Berechnungsstruktur, die sich auf etwas bezieht, d.h. die mit einer Intention verbunden ist. Es ist eine elementare kognitive Einheit, die mit anderen Kognitons zusammen eine dynamische mentale Konfiguration bildet, die sich in der Zeit entwickelt. Kognitons können erzeugt werden oder verschwinden als Folge des Erzeugens oder Verschwindens anderer Kognitons im kognitiven Apparat eines Agenten.

Ein typischer Veränderungsschritt im kognitiven Apparat ist die *Revision von Annahmen*. Ein kognitiver Agent geht bei seinem Planen und Handeln von bestimmten Annahmen aus, muss aber immer bereit sein seine Annahmen auf Grund seiner Wahrnehmung zu revidieren, wenn sie nicht zutreffen. Andere Veränderungsschritte sind z.B. die Veränderung von Zielen oder die Rücknahme von Versprechen.

5.1.3. Typen von Kognitons

Interaktionale Kognitons

Percept Percepte repräsentieren wahrnehmbare Informationen, die durch Wahrnehmung der Umgebung erhalten worden sind. Diese Informationen sind durch Sensoren und Vorverarbeitung in einem speziellen kognitiven Apparat entstanden.

Information Dies ist ein Kogniton, das eine Annahme eines anderen Agenten beschreibt, die durch eine Nachricht übermittelt wurde.

Befehl (oder **Entscheidung**) Ein Befehl ist ein Kogniton, das vom conativen System kommt und die Operationen angibt, die vom interaktionalen System auszuführen, d.h. in Aktionen des Agenten umzusetzen sind, die die Umgebung des Agenten beeinflussen.

Anforderung Durch Anforderungen veranlasst ein Agent einen anderen Agenten zur Ausführung bestimmter Aktionen. Eine Anforderung ist selbst eine Aktion, eine so genannte *relationale Aktion*, weil sie zwischen Agenten stattfindet.

Norm Normen sind Kognitons, die die sozialen Forderungen und die Constraints, die von der Organisation an einen einzelnen Agenten gestellt werden bzw. ihm auferlegt werden. Sie können unterschiedliche Formen haben. Die häufigste ist eine Menge von Aufgaben in Verbindung mit der Rolle, die ein Agent in der Organisation spielt. Solche Normen werden auch als soziale Forderun-

gen bezeichnet. Sie werden in den meisten Fällen nicht vom Agenten erworben, sondern ihm von Entwickler implementiert.

Repräsentationale Kognitons

Annahme Annahmen beschreiben den Zustand der Welt aus der Sicht eines Agenten, d.h. seine Repräsentation der Umgebung, anderer Agenten und seiner selbst. Es ist zu unterscheiden zwischen den Annahmen eines Agenten über sich selbst und über andere. Die Ersteren sollten konsistent mit dem Zustand sein, in dem er sich befindet, während die Letzteren dies nicht sein müssen, sie sind partiell und können sogar falsch sein.

Hypothese Hypothesen sind mögliche Repräsentationen der Welt und anderer Agenten, die von dem Agenten noch nicht „geglaubt“ werden. Eine Hypothese kann eine Aktion auslösen, die zu ihrer Transformation in eine Annahme dient.

Conative Kognitons

Tendenz/Ziel Tendenzen oder Ziele sind im Wesentlichen das Ergebnis von Impulsen und Forderungen. Für die Definition eines Ziels ist es nicht entscheidend, ob es erreicht werden kann, wesentlich ist, dass man sie anstrebt. Ziele bestehen so lange, wie der Grund ihres Bestehens gültig ist. Ziele können durch Anziehung und Abstoßung beschrieben werden.

Trieb Triebe sind interne Ursachen für die Konstruktion von Zielen. Sie entsprechen den Forderungen des vegetativen Systems, z.B. der Forderung den Hunger zu stillen oder zu versuchen, sich selbst zu reproduzieren, allgemein also das zu tun, was das Überleben des Individuums oder der Art sichert.

Forderung Forderungen sind äußere Ursachen für Ziele. Solche sind z.B. Bitten eines anderen Agenten.

Intention Intentionen beschreiben, was einen Agenten motiviert, was es ihm erlaubt eine Handlung auszuführen.

Versprechen Versprechen charakterisieren die Abhängigkeiten, die einen Agenten in Bezug auf sich selbst oder andere Agenten binden, wenn sie sich zur Ausführung einer Aktion oder eines Dienstes entschließen oder wenn sie die Intention haben etwas zu tun.

Organisationale Kognitons

Methode Methoden sind Techniken, Regeln, Pläne und Programme, die ein Agent benutzen kann um eine Aufgabe auszuführen und ein Ziel zu erreichen. Eine Methode beschreibt die einzelnen Schritte, die dazu notwendig sind.

Aufgabe Aufgaben gibt es auf zwei verschiedenen Ebenen, der konzeptuellen und der organisationalen Ebene. Auf der konzeptuellen Ebene beschreibt eine Aufgabe, was getan werden muss um ein Ziel zu erreichen. Die Beschreibung zählt nur ausführbare Aktionen auf, es wird nichts über die Art und Weise ihrer Ausführung gesagt. Dies ist in den Methoden beschrieben. Auf der organisationalen Ebene sind Aufgaben die einzelnen Schritte bei der Ausführung einer Methode.

5.2. Das Interaktionssystem

Ein Agent kann Information über die Welt durch seine Wahrnehmung bekommen. Die Wahrnehmung bildet gewissermaßen eine Tür zwischen der Welt und dem Inneren des Agenten. Der Agent baut eine Repräsentation der Welt auf, in der sein Schlussfolgern und seine Aktionen bedeutsam

sind. Jede Wahrnehmung ist individuell und trägt zu einer persönlichen Repräsentation der Welt bei. Das Wahrnehmungssystem eines Agenten ist mit seinen kognitiven Fähigkeiten und seinen Verhaltensmöglichkeiten verknüpft.

Man unterscheidet zwischen passiven und aktiven perzeptiven Systemen. Diese Unterscheidung beruht auf verschiedenen Auffassungen über die Art der Wahrnehmung. In der philosophischen Tradition seit Aristoteles bis Kant wurde Wahrnehmung als ein passiver Vorgang betrachtet. Danach besteht die Welt aus Objekten und Situationen, die unabhängig vom Betrachter gegeben sind, der Beobachter registriert und klassifiziert nur die Informationen, die von seiner Umgebung kommen. Dementsprechend ist die Bedeutung der Äußerungen eines Sprechers direkt auf die Objekte der Welt bezogen, nicht auf mentale Konstrukte, die unabhängig von diesen Objekten aufgebaut wurden.

Demgegenüber wird bei der Auffassung der Wahrnehmung als aktiver Vorgang angenommen, dass die Wahrnehmung im Gehirn konstruiert wird, ausgehend von den Eingaben der Sinnesorgane und dem bereits existierenden Vorwissen auf Grund früherer Wahrnehmungen und kognitiver Prozesse. Diese Auffassung wird seit Kant vertreten und durch neuere Erkenntnisse aus der Neurophysiologie und Kognitionswissenschaft gestützt. Man kann die beiden Auffassungen von Wahrnehmung schematisch in der BRIC-Notation darstellen, siehe Abbildung 5.1.

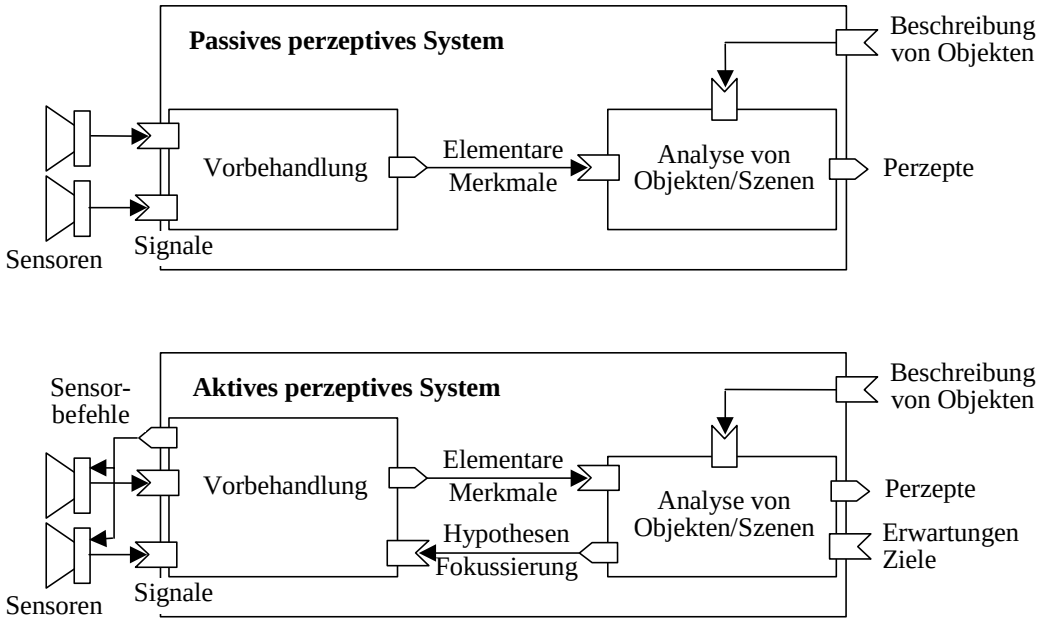

Abbildung 5.1

5.3. Das Repräsentationssystem

5.3.1. Was ist Wissen?

Das Standardverständnis von Wissen in der KI ist, dass darunter alle Informationen zu verstehen sind, die ein Mensch oder ein künstliches System benötigen um eine Aufgabe durchzuführen, die als komplex betrachtet wird.

Der Begriff *Wissen* in diesem Sinn hat wesentlich pragmatischen Charakter. Ein Agent benötigt aber auch Informationen über seine Umgebung und andere Agenten, mit denen er interagiert. Im Allgemeinen unterscheidet man zwischen *etwas wissen* und *wissen, wie man etwas tut*. Das erste betrifft das Verständnis wahrgenommener Dinge und Phänomene. Das zweite betrifft die Analyse

von Beziehungen zwischen Dingen und Phänomenen. Damit kann man die Gesetze des Universums modellieren und die Entwicklung der Dinge vorhersagen.

Der Begriff des Wissens kann am besten durch Angabe von Konzepten und Modellen erklärt werden, die die Phänomene der Entwicklung und des Gebrauchs dessen beschreiben, was es einem Agenten erlaubt, komplexe Aufgaben auszuführen.

Konzeptualisierung und Schlussfolgern

Was sind und vor allem wie entstehen grundlegende Konzepte als die wesentlichen Bestandteile des Wissens? Wenn man Wissen im Sinne von Know-How erklärt, dann kann man nicht über Konzepte reden ohne ihren Gebrauch zu berücksichtigen. Wissen existiert nicht ohne die Probleme, die sich einem Agenten stellen und die die Motivation für den Erwerb von Wissen darstellen, und nicht ohne die Aktionen, die es in die Praxis umsetzen. Die wesentliche Aktion ist das Schlussfolgern, deshalb gehört dieses untrennbar zum Wissen eines Agenten hinzu.

In der traditionellen KI besteht die Vorstellung, dass das ganze Wissen der Menschheit sich auf eine endliche Menge von Konzepten zurückführen lässt und dass es somit möglich ist, eine universelle Wissensbasis zu bauen. Was ist aber ein Konzept? Es gibt keine einfache Definition des Begriffs. Die Erklärung hängt davon ab, was man damit tun will, welcher Aspekt daran einen interessiert.

Ein besonders wichtiger Aspekt des Wissens ist, dass es nicht eine individuelle Angelegenheit ist. Den isolierten Spezialisten gibt es nicht, vielmehr ist das Wissen eines Menschen großenteils das Ergebnis der Interaktion mit anderen Menschen. Dies wird bestätigt durch das Phänomen der Wolfskinder, bei denen sich mangels Interaktion mit Menschen die typisch menschliche Komponente der Intelligenz nicht entwickeln konnte. Aber auch Spezialisten haben ihr Wissen gelernt, wofür der Kontakt zu Lehrern wichtig war und sie praktizieren ihr Wissen zusammen mit anderen Spezialisten auf ähnlichen Gebieten, so dass bei der praktischen Nutzung immer ein Austausch und eine Weiterentwicklung stattfindet. Soziologen haben nachgewiesen, dass für die Entwicklung und Stabilisierung wissenschaftlicher Konzepte soziale Beziehungen wesentlich waren.

Wissen in der Interaktion

Wissen entwickelt sich in der Interaktion mit der Welt und mit anderen Agenten. In der Sozialpsychologie wurde die Bedeutung sozialer Kontakte für die Bildung von Konzepten nachgewiesen. Ein Individuum muss seine Kategorisierungen und Urteile so modifizieren, dass es in Übereinstimmung mit seinen früheren Ansichten und denen der Gruppe, in der es lebt, ist. Wissen ist das Ergebnis von Interaktionen zwischen kognitiven Agenten, die durch Widerspruch, Einwände, Beweise und Widerlegungen Konzepte schaffen, Theorien konstruieren und Gesetze formulieren. Als Slogan der MAS-Entwicklung könnte also formuliert werden:

Es gibt kein Wissen ohne Interaktion.

Hewitt vertritt in seinen Open Information Systems explizit diesen Standpunkt. In ihnen ist Wissen das Ergebnis der Interaktion mehrerer Mikrotheorien. Eine Mikrotheorie kann als das Wissen eines Agenten betrachtet werden. Die einzelnen Mikrotheorien sind kohärent, aber verschiedene Mikrotheorien können in Konflikt geraten oder sogar widersprüchlich sein, obwohl sie aus der selben Organisation stammen. Wenn mehrere Agenten eine gemeinsame Aktion ausführen wollen, müssen sie sich durch Verhandlung darauf einigen, welche Mikrotheorie gelten soll.

Ein wichtiger Aspekt einer Mikrotheorie ist, dass sie einen lokalen Geltungsbereich hat und dass ihre Anwendung relativ zu einem Kontext ist. Zum Beispiel gilt die Mikrotheorie *Etwas, das nicht*

gehalten wird, fällt runter für viele alltägliche Bereiche, aber nicht für alle, z.B. nicht für Ballone oder für interstellare Objekte. Deshalb hat in einem MAS kein Agent vollständiges Wissen über die Welt, sondern immer nur lokales.

Wissen kann auch nicht auf bloßes gelerntes deklaratives Wissen reduziert werden. Bedeutungen können nur aus der Erfahrung erklärt werden, d.h. indem auf das Verhalten eines Agenten in einer Umgebung und unter anderen Agenten referiert wird. Wissen ist also nicht nur „Buchwissen", sondern in starkem Maß Know-How.

5.3.2. Repräsentation von Wissen und Annahmen

Es gibt drei Hauptformen der Wissensrepräsentation:

- symbolisch-kognitivistisch
- konnektionistisch
- interaktionistisch

Die symbolisch-kognitivistische Position lässt sich in drei Grundsätzen zusammenfassen:

(1) Repräsentationen sind unabhängig vom zu Grunde liegenden physikalischen Träger.

(2) Mentale Zustände sind repräsentational oder intentional, d.h. sie beziehen sich auf etwas für die Agenten Externes.

(3) Repräsentationen bestehen aus Symbolen, die in Computer-basierten Sätzen oder irgendwelchen anderen Strukturen angeordnet sind, und Schlussfolgern bedeutet das Manipulieren dieser Strukturen um neue Strukturen aus ihnen zu erhalten.

Nach der konnektionistischen Position ist Wissen in einem Netz wie einem zellulären Automaten in Form numerischer Werte verteilt, die den Verbindungen zwischen den Zellen zugeordnet sind. Schlussfolgern bedeutet Propagieren numerischer Werte in diesem Netz und Veränderung der Werte an den Verbindungen.

Die Wissensrepräsentation nach der konnektionistischen Position heißt auch subsymbolisch, im Kontrast zur symbolischen Repräsentation. Das Ziel dieser Position ist die Vorgänge im Nervensystem von Lebewesen zu simulieren.

Die interaktionistische Position postuliert, dass das Wissen eines Individuums selbst als ein eigenes MAS betrachtet werden kann. Die Konzepte, Ideen und Repräsentationen werden als Agenten spezieller Art (*noetische* Agenten) aufgefasst, die in einem Agenten leben. Schlussfolgern besteht aus den Interaktionen dieser Agenten. Mit der symbolisch-kognitivistischen Position wird die Ansicht geteilt, dass die kognitiven Elemente unabhängig von einem physikalischen Träger sind, und mit der konnektionistischen Position die Ansicht, dass mit den Symbolen keine Bedeutung verbunden ist, dass diese vielmehr durch die Interaktionen zwischen den externen und internen Milieus der Agenten entsteht.

Die symbolisch-kognitivistische Hypothese

Repräsentationen werden in der Form von Symbolen ausgedrückt, die direkt auf die Entitäten der Welt eines Agenten referieren. Die Symbole werden in einer internen Sprache ausgedrückt. Diese ist durch eine formale Grammatik und durch bestimmte Operationen, die auf sie angewandt werden können, definiert. Die Grammatik legt durch Regeln fest, wie die Symbole zu Strukturen zusammengesetzt werden können. Schlussfolgern besteht in der Manipulation der Symbolstrukturen um

neue Symbolstrukturen zu erhalten. Dieser Prozess heißt Inferenz. Die dabei im einzelnen verwendeten Operationen müssen korrekt sein, d.h. sie müssen bestimmte semantische Bedingungen erfüllen und die abgeleiteten Strukturen müssen gültig sein. Diese Art des Schlussfolgerns beruht auf der Existenz folgender fünf Dinge:

(1) einer Menge von Symbolen;

(2) grammatischer Regeln zur Komposition von Symbolen um symbolische Aussagen oder Strukturen zu bilden;

(3) einer Referenzwelt aus klar unterscheidbaren Einheiten, auf die sich die Symbole beziehen können;

(4) einer Interpretationsfunktion, die den Symbolen Bedeutungen gibt, d.h. die jedem Symbol einen Referenten oder eine Klasse von Referenten zuordnet;

(5) einem Inferenzsystem, das die Ableitung neuer Aussagen oder Strukturen ermöglicht.

5.3.3. Logik des Lernens und der Annahmen

Schlussfolgern beruht immer auf Annahmen. Diese Annahmen können nur durch Interaktion mit anderen Entitäten verifiziert werden (Personen, Anzeigen), deren Zuverlässigkeit ebenfalls nur angenommen werden kann. Das heißt, Fakten als solche gibt es nicht, es gibt nur Annahmen, die noch nicht widerlegt wurden.

Zu den Annahmen gehört wesentlich das Prinzip der Revision. Jede Information und jede Annahme kann in Frage gestellt werden. Die Fähigkeit Fakten, eine Deduktion, ein Gesetz oder ein Urteil in Frage zu stellen ist die Grundlage der menschlichen Fähigkeit der kognitiven Adaption, d.h. der Fähigkeit unser kognitives System an die sich ständig entwickelnde Welt anzupassen. Diese ständige Entwicklung ist eine Folge der Interaktionen verschiedener Agenten in und mit der Welt.

Eine weitere wichtige Eigenschaft von Annahmen ist, dass sie nicht universal sind. Sie sind vielmehr relativ zu einem Individuum, einer Gruppe oder einer Gesellschaft. Die Annahmen verschiedener Individuen in einer Gruppe müssen nicht kohärent sein. Manche Individuen haben bessere oder neuere Annahmen als andere. Für MAS ist das ein Hinweis auf die Notwendigkeit lokaler Datenhaltung und von Kommunikation zum Austausch und Vergleich von Annahmen.

Formale Definition von Annahmen

Die epistemische Logik befasst sich mit der Repräsentation von Wissen und Annahmen. Sie ist eine Erweiterung der Logik erster Stufe. Es werden die Prädikate `know` und `believe` benutzt und sie werden in eine bestimmte Beziehung zueinander gesetzt und erhalten eine besondere Interpretation. Die Beziehung zwischen beiden ist in folgender Weise definiert:

$$\texttt{know(A,P)} = \texttt{believe(A,P)} \wedge \texttt{true(P)}$$

Das bedeutet, was ein Agent weiß, ist auf jeden Fall wahr. Man kann also folgern

$$\texttt{know(A,P)} \Rightarrow \texttt{P}$$

Mögliche Welten

Jeder Agent kann sich eine Anzahl von Welten vorstellen, die er für denkbar hält. Dann glaubt ein Agent eine Aussage, wenn sie in allen Welten, die er sich vorstellt (d.h. in allen für ihn möglichen Welten), wahr ist.

Auf der Basis der möglichen Welten wird die Interpretation eines Satzes mit Hilfe von Kripke-Modellen definiert. Ist $W = \{w_1, ..., w_k\}$ eine Menge möglicher Welten und S eine Menge von Sätzen, dann ist das n+2-Tupel

$$M = \langle W, I, R_1, ..., R_n \rangle$$

ein Kripke-Modell. Dabei ist R_i die Zugänglichkeitsrelation des Agenten i und I eine Bewertungsfunktion, die Paaren aus Sätzen und Welten Wahrheitswerte zuordnet, d.h. eine Funktion der folgenden Art:

$$I: S \times W \rightarrow \{True, False\}$$

Die Interpretation eines Satzes φ, der eine Annahme enthält, wird nun durch die Relation *semantische Konsequenz* definiert. Sie wird notiert durch

$$M, w \models_I \varphi$$

Dies bedeutet: φ ist wahr in der Welt w des Modells M. Für beliebige Sätze wird diese Definition entsprechend ihrem syntaktischen Aufbau rekursiv fortgesetzt:

$M, w \models_I \varphi$	gdw. φ ein atomarer Satz ist und $I(\varphi, w) = True$
$M, w \models_I \neg\varphi$	gdw. $M, w \not\models_I \varphi$
$M, w \models_I \varphi \wedge \psi$	gdw. $M, w \models_I \varphi$ und $M, w \models_I \psi$
$M, w \models_I \varphi \vee \psi$	gdw. $M, w \models_I \varphi$ oder $M, w \models_I \psi$
$M, w \models_I \varphi \Rightarrow \psi$	gdw. $M, w \not\models_I \varphi$ oder $M, w \models_I \psi$
$M, w \models_I$ `believe`(a, φ)	gdw. für jede Welt w_i mit $w\ R_a\ w_i$ gilt $M, w_i \models_I \varphi$

Für den Modaloperator `believe` werden verschiedene Eigenschaften durch Axiome definiert. Diese sind das Distributionsaxiom, das Axiom der Widerspruchsfreiheit, die Axiome der positiven und negativen Introspektion und das Lernaxiom. Das Distributionsaxiom ist eine Anwendung des Modus Ponens. Es gibt verschiedene äquivalente Formulierungen, u.a. die folgenden beiden:

```
believe(A,(P ⇒ Q)) ⇒ believe(A,P ⇒ believe(A,Q))          (K)
believe(A,P) ∧ believe(A,(P ⇒ Q)) ⇒ believe(A,Q)
```

Das Axiom der Widerspruchsfreiheit besagt, dass ein Agent nicht gleichzeitig an einen Sachverhalt und an sein Gegenteil glaubt.

```
believe(A,P) ⇒ ¬believe(A,¬P)                            (D)
```

Die Introspektionsaxiome stellen eine Art positive bzw. negative Vergewisserung von Agenten über ihre Annahmen dar.

```
believe(A,P) ⇒ believe(A,believe(A,P))                   (4)
¬believe(A,P) ⇒ believe(A,¬believe(A,P))                 (5)
```

Das Lernaxiom definiert von Agenten angenommene Sachverhalte als wahr:

```
believe(A,P) ⇒ P                                         (T)
```

Mittels der Logik der möglichen Welten lassen sich auch wechselseitige Annahmen zwischen verschiedenen Agenten formulieren, wie die folgenden Beispielsätze illustrieren.

```
believe(Hans,voll(tank(Tina)))
believe(Eva,leer(tank(Tina)))
believe(Eva,believe(Hans,leer(tank(Tina))))
```

Die Sätze bedeuten, dass in allen Welten, die zu Beginn, d.h. von w_0 aus, für Hans zugänglich sind, der Tank von Tina voll ist, während in allen Welten, die zu Beginn für Eva zugänglich sind, der Tank leer ist. Der letzte Satz besagt, dass in allen für Eva zugänglichen Welten der Satz

```
believe(Hans,leer(tank(Tina)))
```

wahr ist und das heißt, dass in allen Welten, die für Hans zugänglich sind von den Welten aus, die für Eva zu Beginn zugänglich sind, der Satz

```
leer(tank(Tina))
```

gilt. Dieser Zusammenhang ist in Abbildung 5.2 illustriert.

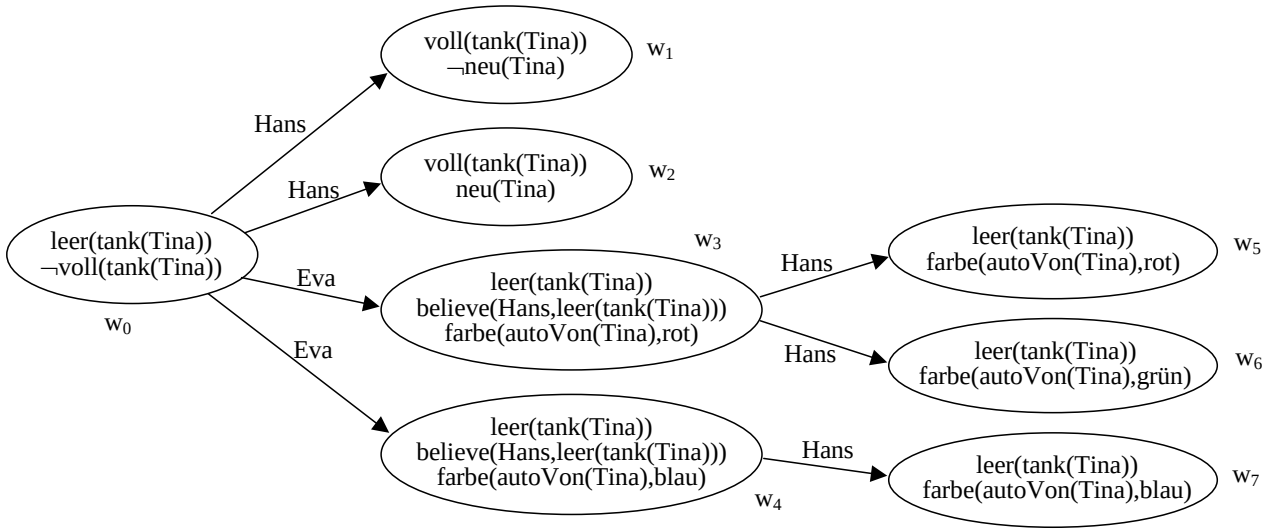

Abbildung 5.2

Operationale Logik

Die *operationale Logik* oder *Satzsemantik* ist ebenfalls eine Logik der Annahmen, sie kommt aber ohne mögliche Welten aus, vielmehr ist die Semantik eines Ausdrucks durch die formale Manipulation von Sätzen definiert.

Jedem Agenten *a* ist eine Wissensbasis Δ_a, bestehend aus gewöhnlichen logischen Sätzen, und eine Menge Γ_a aus Inferenzregeln zugeordnet. Die Sätze der Wissensbasis zusammen mit den aus ihnen durch Anwendung der Inferenzregeln ableitbaren Sätze bilden die Theorie Θ_a des Agenten *a*. Mit der Notation $\Delta_a \vdash_a \varphi$ wird beschrieben, dass der Satz φ wahr ist in der Theorie Θ_a, d.h. dass φ ein aus der Wissensbasis des Agenten *a* abgeleitetes Theorem ist. Ein Agent glaubt einen Satz φ, wenn $\varphi \in \Theta_a$, d.h. das, was er glaubt oder annimmt, entspricht seinen deduktiven Fähigkeiten. Will man z.B. nachweisen, dass eine Agent *a* den Satz $\varphi \Rightarrow \psi$ glaubt, dann muss man aus den Sätzen der Wissensbasis und mit den Inferenzregeln des Agenten beweisen, dass $\varphi \vdash_a \psi$ gilt.

5.3.4. Adäquatheit und Revision von Annahmen

Ein Modell ist adäquat, wenn es einen Morphismus zwischen der Struktur des Modells und der Struktur der realen Welt gibt, ansonsten ist es inadäquat.

Inadäquatheit der Modelle führt manchmal zu Schwierigkeiten bei Handlungen. Zum Beispiel glaubten am Ende des 19. Jahrhunderts viele Leute, dass Maschinen, die schwerer als Luft sind, nicht fliegen könnten. Das Modell dieser Leute beruhte nur auf dem Konzept *Gewicht*, schloss aber das Konzept der *Aerodynamik* aus und war deshalb inadäquat.

Ein Modell besteht aus Annahmen, d.h. Daten, die in Frage gestellt werden können, so bald neue Informationen zur Verfügung stehen. Deshalb muss man in der Lage sein, das was man annimmt bei Bedarf zu revidieren.

Default-Logik

In der realen Welt sind viele Aussagen nicht unbegrenzt oder universell gültig, vielmehr ist ihre Gültigkeit relativ zum aktuellen Zustand der Welt. In der klassischen Logik kann man diesen Sachverhalt nicht adäquat wiedergeben. Sobald in ihr eine Aussage gemacht wird, kann sie nicht mehr in Frage gestellt werden, sie ist ab sofort unveränderlich gültig. Deshalb heißt die klassische Logik auch *monoton*, d.h. es werden beim Schlussfolgern nur immer neue Fakten zur Wissensbasis hinzugefügt. Die Wahrheitswerte der Sätze in der Wissensbasis können nicht verändert werden ohne die ganze Wissensbasis inkonsistent zu machen.

Da die reale Welt sich entwickelt und da man praktisch immer nur partielles Wissen über die Welt haben kann, braucht man geeignete Logiken, die es erlauben, Wissen zu revidieren und damit Fakten in Frage zu stellen, die anfangs abgeleitet wurden, aber inzwischen nicht mehr gelten. Es muss möglich sein, einzelne Sätze zu verändern oder zu entfernen, ohne gleich die ganze Wissensbasis zu verwerfen, um so der inkrementellen und progressiven Art, in der sich die Annahmen eines Agenten entwickeln, gerecht zu werden. Die Daten, mit denen ein Agent normalerweise umgeht, sind unvollständig, unsicher und der Entwicklung unterworfen.

Die theoretische Grundlage für revidierbare Logiken bilden die *nichtmonotonen* Logiken, die als Erweiterungen der klassischen Logik betrachtet werden. Ursprünglich wurden sie für einen anderen Zweck entwickelt, nämlich um bestimmte Aspekte des Common Sense Reasoning zu formalisieren, insbesondere solche, die sich mit *typischen* Aussagen beschäftigen, also Aussagen der Form „Im Allgemeinen sind alle A‘s auch B‘s“.

Die Sätze der Default-Logik haben die selbe Form wie die der Logik erster Stufe. Default-Aussagen werden durch spezielle Inferenzregeln, genannt *Default-Regeln*, formuliert. Ihr allgemeine Form ist

$$\frac{\alpha : M\beta}{\gamma}$$

Die Bedeutung der Regel ist: Wenn α eine Annahme ist und der Satz β nicht inkohärent mit den anderen ist, dann kann γ angenommen werden. Sehr häufig sind β und γ der selbe Satz, diese Regeln nennt man *normale* Defaults.

Mit Default-Regeln kann man auch das revidierbare Wissen eines Agenten formulieren. Zum Beispiel lässt sich die Aussage, dass ein Roboter mit einem Greifer normalerweise ein Objekt ergreifen kann, durch folgende Default-Regel wiedergeben:

$$\frac{\texttt{greiferFrei(x)}: M\ \texttt{kannErgreifenObjekt(x)}}{\texttt{kannErgreifenObjekt(x)}} \qquad \text{(RD1)}$$

$$\frac{\texttt{tankuhr(x,voll)}: M\ \texttt{tankVoll(x)}}{\texttt{tankVoll(x)}} \qquad \text{(RD2)}$$

Bei einer Default-Regel ist die Verbindung zwischen der Prämisse und der Konklusion lockerer als in der Logik erster Stufe. Besteht ein Widerspruch zwischen der Konklusion und anderen Fakten der Wissensbasis, dann ist die Regel nicht anwendbar. Das heißt, *vor* jeder Anwendung der Regel muss geprüft werden, ob eventuell ein solcher Widerspruch vorliegt.

Rechtfertigungen und das TMS

Die Default-Logik kann relativ leicht implementiert werden mit Hilfe von Truth Maintenance Systemen (TMS). Mit ihnen kann man Rechtfertigungen oder Begründungen für Annahmen behandeln. Sie beschreiben Abhängigkeiten zwischen Fakten der Wissensbasis und drücken aus, ob ein Fakt geglaubt werden kann oder nicht, indem sie positive und negative Rechtfertigungen für das Fakt angeben. Ein *TMS* ist also ein Netz von Abhängigkeiten zwischen Daten, die Knoten des Netzes repräsentieren Daten, die wahr oder falsch sein können, und die Kanten charakterisieren die Abhängigkeiten zwischen den Daten. Die Kanten können *positiv* oder *negativ* sein. Positive Kanten repräsentieren positive Rechtfertigungen zwischen den Daten und negative Kanten negative Rechtfertigungen. Ein Knoten heißt *gültig*, wenn alle Knoten, von denen er positiv abhängt, gültig sind, und alle Knoten, von denen er negativ abhängt, ungültig sind. Abbildung 5.3 zeigt ein Beispiel für die Aussage, wann Rudi ein Objekt ergreifen kann.

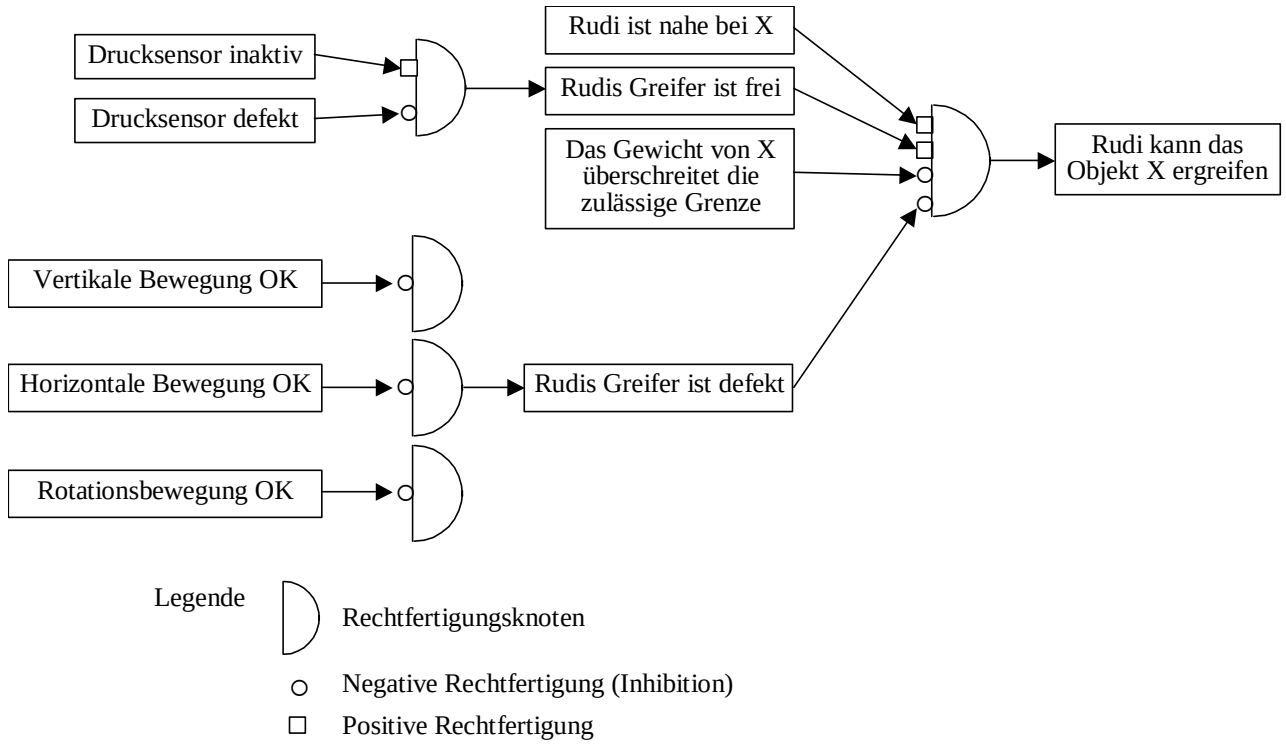

Abbildung 5.3

Die Daten in einem TMS sind entweder primitiv oder von anderen Daten abhängig. Jedem nicht primitiven Datum ist gewissermaßen ein kleiner Prozessor zugeordnet, der die Gültigkeit der Aussagen, von denen das Datum abhängt, berechnet. Knoten, die keine Rechtfertigung haben, sind entweder *Prämissen* oder *Hypothesen*. Prämissen sind solche Knoten, deren Daten als immer gültig betrachtet werden, Hypothesen sind Knoten, von deren Daten *angenommen* wird, dass sie gültig sind. Diese Knoten können sich beim Arbeiten mit dem TMS als ungültig erweisen.

Vorhersagen und Adäquatheit

Ein Vorhersagesystem kann auf der Basis von Wahrnehmungen oder von Daten über den aktuellen Weltzustand, auf der Basis von Modellen über die von einem Agenten ausführbaren Aktionen und auf der Basis der Gesetze der Welt Erwartungen erzeugen, d.h. Perzepte oder Daten, die durch die vom Wahrnehmungssystem gelieferten Perzepte und Daten bestätigt werden. Die Vorhersageprozedur kann formal wie folgt beschrieben werden. Sei P_1 eine Menge von Perzepten oder Daten, die einen Teil das aktuellen Zustands σ_1 beschreiben, sei $M(Op)$ ein Modell der Aktion, die Agent a ausführt, sei L die Menge der Gesetze des Universums und C die Menge der gemachten Annahmen. Dann ist

$$P_1 = Perzept_a(\sigma_1)$$
$$P_2' = Vorhersage_a(P_1, M(Op), L, C)$$

wobei P_2' die Menge der erwarteten Perzepte oder Daten ist, die im Zustand σ_2 wahrgenommen werden müssten. Nach dem Modell der Aktion als Produktion von Einflüssen (Abschnitt 4.3) lässt sich σ_2 folgendermaßen berechnen:

$$\sigma_2 = React(\sigma_1, Exec(Op, \sigma_1) \,||\, \gamma_{andere})$$

wobei γ_{andere} die von anderen Agenten verursachten Einflüsse sind. Der Zustand σ_2 kann aber von Agent a in Form einer Menge von Perzepten wahrgenommen werden:

$$P_2 = Perzept_a(\sigma_2)$$

Stimmen P_2 und P_2' überein, d.h. sind nicht inkohärent, dann ist das Vorhersagemodell adäquat. Weichen P_2 und P_2' stark voneinander ab oder sind sogar widersprüchlich, dann müssen im Vorhersagemodell die Repräsentationen revidiert werden, die für die Abweichung oder den Widerspruch verantwortlich sind.

Für die Revision des Vorhersagemodells benötigt man eine Revisionsstrategie. Mit dieser werden die Elemente des Modells festgestellt, die als Verursacher der Diskrepanz am stärksten in Frage kommen. Dafür gibt es verschiedene Standardmöglichkeiten:

(1) *Die Sensoren arbeiten fehlerhaft*. Dies kann durch direkte Überprüfung der Sensoren oder durch Vergleich mit den Wahrnehmungen anderer Agenten festgestellt werden. Man kann davon ausgehen, dass die Sensoren mehrerer Agenten nicht alle gleichzeitig fehlerhaft sind.

(2) *Die von anderen Agenten erhaltenen Daten sind nicht mehr korrekt oder sind unzureichend*. In diesem Fall können die Daten erneut abgefragt werden, falls dies möglich ist, oder man kann versuchen, ergänzende Daten zu bekommen.

(3) *Die Annahmen des Agenten sind falsch oder unzureichend*. Dies zu überprüfen kann sehr aufwändig sein, da die Annahmen von einer großen Zahl von Perzepten, Daten und Gesetzen der Welt abhängen können. Diese alle zu prüfen ist im Allgemeinen nicht möglich. Man kann statt dessen versuchen die verdächtigsten Elemente festzustellen und zu revidieren und erst im Bedarfsfall auf die weniger verdächtigen zurück zu greifen. Informationen von anderen Agenten können ebenfalls hilfreich sein.

(4) *Die Aktion entspricht nicht seinem Modell*. In diesem Fall gibt es zwei Möglichkeiten. Die Ursache kann eine falsche Modellierung der Aktion sein. Dann muss das Modell revidiert und die Aktion unter anderen Bedingungen getestet werden. Die Ursache kann aber auch ein Mangel

an Fähigkeit zur Ausführung der Aktion sein. In diesem Fall kann die Diskrepanz nur durch Verbesserung der Fähigkeit, d.h. durch Üben, behoben werden.

(5) *Die Gesetze des Universums oder die psychologischen Modelle der anderen Agenten sind nicht mehr adäquat oder erweisen sich als unzureichend*. Die Gesetze des Universums müssen durch Experimente nachgeprüft werden um ihre Gültigkeit zu verifizieren. Dieser Vorgang kann immer wieder erforderlich sein, wenn im Licht neuer Erkenntnisse die bisherigen Modelle fraglich werden. Durch Methoden des Maschinellen Lernens oder durch Interaktion zwischen Agenten, insbesondere zwischen Lehrer und Schüler, kann neues Wissen erworben und altes aktualisiert werden.

(6) Vorhersagefehler können in der Komplexität der Situation begründet sein, d.h. in der Tatsache, dass es zu viele Parameter gibt und dass es deshalb fast unmöglich ist eine Vorhersage zu machen.

5.4. Inhalte von Repräsentationen

Modelle können in zwei Klassen unterteilt werden: *Faktische Modelle* und *provisorische Modelle*. Faktische Modelle enthalten die Annahmen eines Agenten über den gegenwärtigen Zustand der Welt und über Agenten (einschließlich seiner selbst), provisorische Modelle enthalten Gesetze und Theorien, die die Antizipation zukünftiger Weltzustände und zukünftigen Verhaltens von Agenten erlauben. Beide Klassen von Modellen können entsprechend der Dimensionen der Analyse von Organisationen (Abschnitt 3.2.2) verschiedene Formen annehmen. Annahmen können nach den vier Dimensionen sozial (σ), relational (ρ), Umgebungs- (χ) und personal (π) analysiert werden. Die faktischen Annahmen eines Agenten sind aus Kognitons gebildet, die aus Wahrnehmungen und früheren Annahmen stammen. Wahrnehmungen und Informationen sind die wichtigsten Quellen für den Aufbau faktischer Annahmen. Erstere dienen zum Aktualisieren von Annahmen über die physikalische Natur von Objekten und der Welt, Letztere ermöglichen eine Repräsentation der Welt und anderer Agenten ohne sie wahr zu nehmen oder in ihrer Nähe zu sein. Aus den so entstandenen Annahmen kann ein Agent neue durch Schlussfolgern ableiten. Provisorische Modelle werden vom Entwickler implementiert oder durch Lernen erworben.

5.4.1. Umgebungsannahmen (χ)

Umgebungsannahmen beziehen sich auf alles, was beobachtet oder wahrgenommen werden kann. Dazu gehört der Zustand der Umgebung und die wahrnehmbaren Eigenschaften von Objekten und Agenten. Letztere unterscheiden sich unter diesem Aspekt nicht von Objekten, weil hier nur äußere Charakteristiken wie Position, Form oder Farbe von Interesse sind.

Umgebungsannahmen können fehlerhaft sein und zu falschen Interpretationen und abwegigen Entscheidungen führen. Deshalb ist es wichtig, sie mit Revisionsmechanismen zu versehen, die die natürliche Entwicklung dieser Daten berücksichtigen.

Auch die Gesetze des Universums können in der Umgebungsdimension repräsentiert werden. Ist dies der Fall, dann hat der Agent echte provisorische Modelle zur Verfügung, denn er kann damit versuchen, sich selbst in die Zukunft zu projizieren und den zukünftigen Zustand der Welt zu antizipieren.

5.4.2. Soziale Annahmen (σ)

Die sozialen Annahmen beziehen sich auf die Rolle und die Funktionen, die in einer Gesellschaft ausgeübt werden, und auf die Organisation und Struktur der Gruppe eines Agenten. Diese Annah-

men haben deshalb einen ähnlichen Allgemeinheitscharakter wie die Umgebungsannahmen, nur dass sie auf die Gesellschaft und nicht die Welt gerichtet sind. Zu den sozialen Annahmen gehören auch alle Standards, Constraints und sozialen Gesetze, die für alle Mitglieder der Gesellschaft gelten.

5.4.3. Relationale Annahmen (ρ)

Die relationalen Annahmen betreffen die Repräsentation von Agenten durch einen Agenten. Sie umfassen die Fertigkeiten, Intentionen, Verpflichtungen, Pläne, Methoden, Aktivitäten, Rollen und Verhaltensweisen der anderen Agenten. Mit Hilfe dieser Annahmen kann ein Agent die anderen Agenten in seine Schlussfolgerungen einbeziehen ohne sie um Informationen bitten zu müssen, und er kann so partielle Pläne erstellen und den zukünftigen Zustand des gesamten MAS antizipieren. Ein Agent macht sich somit ein Bild über andere Agenten und ordnet ihnen u.U. Eigenschaften zu, die sie gar nicht haben. Gerade dadurch ist es aber möglich, dass heterogene Agenten koexistieren können.

Fertigkeiten

Das Wissen über die Fertigkeiten (oder Fähigkeiten) anderer Agenten wird dazu genutzt Arbeit zu verteilen ohne die Agenten erst nach ihren Fertigkeiten fragen zu müssen. Muss ein Agent eine Aufgabe ausführen, für die er nicht die richtigen Fertigkeiten hat, und kennt er einen anderen Agenten, der die notwendige Fertigkeit hat, dann kann er diesen um die Ausführung bitten.

Fertigkeiten können durch Schlüsselwörter, durch atomare Sätze oder Objekte repräsentiert werden. Die Repräsentation durch Schlüsselwörter setzt voraus, dass es ein Vokabular der Schlüsselwörter gibt und dass jeder Agent darüber verfügt.

Für eine genauere Beschreibung von Fertigkeiten reicht diese Repräsentation nicht aus. Deshalb wird häufig eine Darstellung durch atomare Sätze der Logik erster Stufe verwendet. Damit ist es möglich Fertigkeiten wie die Herstellung einer bestimmten Beziehung zwischen verschiedenen Objekten zu beschreiben.

Die Darstellung hat aber den Nachteil, dass man mit ihr nicht die Verbindungen zwischen verschiedenen Fertigkeiten adäquat beschreiben kann. Zum Beispiel ist symbolisches Integrieren eine spezielle Form des symbolischen Rechnens und hat viele Merkmale der allgemeineren Fähigkeit. Um solche Zusammenhänge ausdrücken zu können, werden Objekte benutzt. Jede Fertigkeit wird durch ein Objekt beschrieben und diese lassen sich bezüglich ihres Allgemeinheitsgrades in einer Hierarchie anordnen.

Intentionen und Verpflichtungen

Intentionen und Ziele beschreiben, was Agenten vorhaben. Intentionen führen zu Verpflichtungen sich selbst gegenüber, deshalb genügt es Verpflichtungen von Agenten gegenüber anderen Agenten und sich selbst zu repräsentieren. Kennen Agenten die Verpflichtungen, die andere Agenten eingegangen sind, dann können sie dies bei ihrer eigenen Planung berücksichtigen.

Pläne/Methoden

Kennt ein Agent die Pläne anderer, dann kann er diese in seine eigene Planung einbeziehen.

Aktivitäten

Kennt ein Agent die aktuellen Aktivitäten anderer Agenten, d.h. ihre momentane Auslastung, dann kann er dies berücksichtigen, wenn er versucht Arbeit zu delegieren.

Rollen

Die Rollen betreffen die Aktivitäten, die von den Agenten innerhalb einer Organisation erwartet werden. Wissen über Rollen anderer Agenten kann in ähnlicher Weise wie Wissen über Fertigkeiten beim Delegieren von Arbeit genutzt werden.

Verhaltensweisen

Wissen über Verhaltensweisen anderer Agenten ermöglicht es einem Agenten das zukünftige Verhalten der Agenten vorher zu sagen. Dabei ordnet er den anderen Agenten Intentionalität zu unabhängig davon, ob diese kognitive oder reaktive Agenten sind. Das heißt ob sein Modell der anderen Agenten zutrifft oder nicht spielt keine Rolle, so lange es zutreffende Vorhersagen macht.

5.4.4. Personale Annahmen (π)

Personale Annahmen sind relationale Annahmen eines Agenten über sich selbst. Dem entsprechend gibt es die selben Typen wie bei den relationalen Annahmen. Der Unterschied zu den relationalen Annahmen liegt darin, dass die personalen Annahmen ständig aktualisiert werden können, wie in einem dauernden Lernprozess. Ein Agent repräsentiert also sich selbst als ob er ein anderer wäre, d.h. er objektiviert sich selbst. Das Wissen, das er über sich selbst hat, kann deshalb leicht mit anderem Wissen, etwa über die Welt und andere Agenten, kombiniert werden um Schlussfolgerungen zu ziehen.

5.5. Das conative System

5.5.1. Rationalität und Überleben

Der Begriff der *Rationalität* umfasst das Entwickeln von Zielen, die Abwägung über Ziele und die Auswahl von Aktionen zum Erreichen von Zielen. Ein Agent ist also rational, wenn er die ihm zur Verfügung stehenden Mittel in Hinsicht auf die Ziele einsetzt, die er sich selber gesetzt hat. Formal lässt sich dies so formulieren:

$$\texttt{intention(a,P)} \wedge \texttt{kannTun(a,P)} \Rightarrow \texttt{später(Execute(a,P))}$$

Dem Überleben liegt der Begriff der *Lebensfähigkeit* zu Grunde. Er lässt sich auf MAS mit künstlichen Agenten übertragen, wenn man ihn so versteht, dass er die Fähigkeit eines dynamischen Systems beschreibt die Werte seiner relevanten Parameter innerhalb bestimmter Toleranzbereiche zu halten. Das Verhalten eines lebensfähigen Agenten besteht also darin, seine gesamte Konfiguration zu jeder Zeit so zu gestalten, dass er in dem Bereich der Lebensfähigkeit bleibt.

Verlässt ein Agent den Bereich der Lebensfähigkeit, z.B. wenn er es versäumt, sich rechtzeitig wieder mit Energie aufzuladen, dann stirbt er. Er sollte also versuchen in diesem Bereich zu bleiben, z.B. indem er sich ernährt, vor Jägern flieht, Unfälle vermeidet und eventuell die Aufgaben ausführt, für die er entworfen wurde.

5.5.2. Ein Modell des conativen Systems

Das conative System bestimmt die Aktion, die ein Agent auf der Grundlage der Daten und des Wissens, die ihm zur Verfügung stehen, ausführen sollte um eine Funktion auszuüben, wobei er gleichzeitig seine strukturelle Integrität und seine Aktionsmöglichkeiten in der Weise erhalten sollte, dass die Organisation, zu der er gehört, ebenfalls ihre strukturelle Integrität erhält. Die wesentlichen Konzepte des conativen Systems sind *Tendenzen*, *Motivationen* und *Befehle*.

Alles Verhalten ist das Ergebnis einer Kombination und Interaktion von Tendenzen. Sie sind der Antrieb mentaler Aktivität und des Übergangs zum Handeln. Sie sind Ausdruck eines Mangels, den der Agent zu beheben versucht. Die Mängel sind nicht nur individuell, sondern auch inter-individuell und sozial. Ein gesättigter Agent, d.h. einer ohne Mängel, ist inaktiv. Allgemein bedeutet Sättigung Immobilität, während Mangel an Sättigung zu zielgerichteter Bewegung führt.

Abgesehen von den Tendenzen, die einen Agenten zum Handeln treiben, gibt es auch Tendenzen, die seine Aktionen begrenzen. Zum Beispiel beschränken Tabus und soziale Normen das Repertoire an Aktionen, die als akzeptabel betrachtet werden. Aber auch diese Tendenzen können als Ausdruck eines Mangels betrachtet werden, oder genauer gesagt, der Antizipation eines Mangels, der in der Zukunft eintreten könnte, wenn die Beschränkung der Aktionsmöglichkeit nicht beachtet wird.

Insgesamt lässt sich eine Tendenz wie folgt definieren:

Definition *Tendenz*
Eine *Tendenz* ist das Kogniton, das einen Agenten zum Handeln antreibt, das sein Handeln beschränkt oder das ihn am Handeln hindert.

Tendenzen sind ihrerseits Kombinationen von *Motivationen*, d.h. noch elementareren Kognitons (Perzepte, Triebe, Standards, Verpflichtungen, soziale und inter-personelle Forderungen). Sie entstehen und werden behandelt im motivationalen Subsystem des conativen Systems.

Das Entscheidungs-Subsystem des conativen Systems wertet die Tendenzen in Abhängigkeit von Annahmen, Verpflichtungen, Plänen und Restriktionen aus, wählt die mit der höchsten Priorität aus und bestimmt die Mittel, die zu ihrer Realisierung erforderlich sind. Die Entscheidungsfindung hängt von der Persönlichkeit der Agenten ab und führt zur Definition dynamischer Prioritäten zwischen den Tendenzen. Das Ergebnis der Überlegung ist eine Menge von Entscheidungen, die dem organisationalen System zur Ausführung übergeben werden. Speziell bei kognitiven Agenten führt die Übermittlung einer Entscheidung zur Erstellung eines Plans durch Zerlegen des Ziels in Teilziele. Abbildung 5.4 zeigt die allgemeine Struktur des conativen Systems.

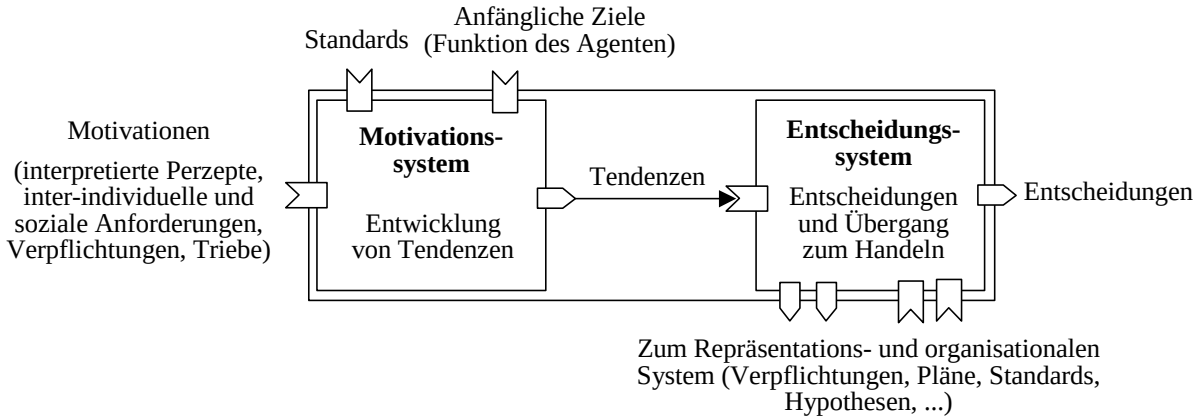

Abbildung 5.4

5.6. Motivationen: Quellen der Aktion

Tendenzen entstehen aus *Motivationen*. Es gibt zahlreiche Motivationen. Im Folgenden werden sie nach den vier Analysedimensionen personal, Umgebungs-, sozial und relational in vier Kategorien gruppiert. Die *personalen Motivationen* richten sich auf den Agenten. Sie zielen auf die Erfüllung

seiner Bedürfnisse oder das Auflösen von Verpflichtungen, die er sich auferlegt hat, ab. *Umgebungsmotivationen* werden durch die Wahrnehmungen des Agenten erzeugt. *Soziale* oder *deontische Motivationen* beziehen sich auf die Forderungen durch die Organisation auf höherer Ebene oder den Entwickler. *Relationale Motivationen* hängen von den Tendenzen anderer Agenten ab.

Eine besondere Quelle von Motivationen stellen Verpflichtungen dar. Sie sind ähnlich grundlegend für die Entstehung von Tendenzen wie Hunger oder Sexualtrieb. Die Verpflichtungen werden durch das organisationale System übergeben und durch das Motivationssystem in Tendenzen umgesetzt. Die aus Verpflichtungen entstandenen Motivationen heißen kontraktuale Motivationen. Es gibt sie in der personalen, sozialen und relationalen Dimension. Ein Agent hat die Freiheit, eine Verpflichtung einzuhalten oder nicht, denn diese werden ja in Tendenzen umgesetzt, über die er in seinem Entscheidungssystem entscheidet.

5.6.1. Personale Motivationen: Vergnügen und Restriktionen

Es gibt zwei Gruppen personaler Motivationen: *Hedonistische Motivationen*, die zum Vergnügen des Agenten beitragen, und *personale kontraktuale Motivationen*, die den Agenten an seine gegen ihn selbst eingegangenen Verpflichtungen erinnern und damit die Persistenz seiner eigenen Ziele sichern.

Als *Triebe* werden die internen Stimuli bezeichnet, die das vegetative System produziert und von denen sich ein Agent nicht befreien kann. Eine Nicht-Befriedigung der Triebe kann unliebsame Folgen für den Agenten und seine Nachkommen haben.

Triebe sind mentale Repräsentationen von Erregungen, die aus dem vegetativen System kommen. Sie werden in Tendenzen umgeformt unter dem Einfluss von Wahrnehmungen, verfügbaren Daten und Annahmen. Sie sind die Grundlage für individuelle Aktionen. Sie stehen in Konflikt mit kontraktualen Motivationen und auch untereinander, weil die von ihnen erzeugten Tendenzen und Ziele oft widersprüchlich sind.

Es ist Teil der *Persönlichkeit* eines Agenten, ob er langfristige Tendenzen oder kurzfristige Tendenzen bevorzugt. Ein sehr prinzipienfester Agent wird Tendenzen bevorzugen, die eine längerfristige Befriedigung garantieren, während ein schwacher Agent Tendenzen bevorzugen wird, die seine Triebe unmittelbar befriedigen, selbst wenn das Ergebnis längerfristig weniger wünschenswert ist.

5.6.2. Umgebungsmotivationen: Verlangen nach einem Objekt

Wahrnehmungen sind die wesentlichen Motivationen reflexartiger Aktionen. Ein Reflex ist eine unmittelbare Reaktion auf einen äußeren Stimulus. Das conative System hat bei Reflexen nur die Aufgabe die Perzepte in Übereinstimmung mit vorprogrammierten Aktionen zu bringen, die Tendenzen sind mit den Perzepten vermischt. Das Entscheidungs-Subsystem filtert nur die Perzepte und stößt die zugeordneten Aktionen an.

Nicht rein reflexartigen Aktionen werden durch eine Kombination von Perzepten und Trieben angestoßen. Beide verstärken sich gegenseitig. Es gibt ein *spezifisches Aktionspotential*, dessen Höhe dafür entscheidend ist, ob eine Aktion oder eine festes Aktionsmuster ausgelöst wird. Dieser Begriff stammt aus der Verhaltensforschung. Konrad Lorenz hat eine Aktivierungsmodell entwickelt, das in Abbildung 5.5 dargestellt ist.

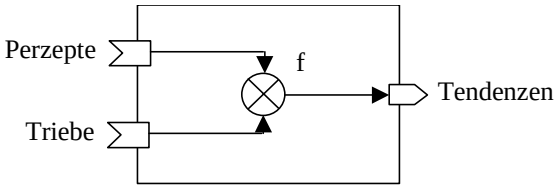

Abbildung 5.5

5.6.3. Soziale Motivationen: Das Gewicht der Gesellschaft

Man kann zwei Typen sozialer Motivationen unterscheiden: funktionale Anforderungen und deontische Regeln. Funktionale Anforderungen sind die Triebkraft der meisten Aktionen, die in Verbindung mit (beruflicher) Arbeit ausgeführt werden. Die Agenten (Menschen) erfüllen ihre Funktionen in der Gesellschaft entsprechend ihren sozialen Positionen und spielen somit ihre Rollen und führen die damit verbundenen Aufgaben aus.

In menschlichen Gesellschaften sind diese Motivationen mit hedonistischen Ursachen verknüpft. Dabei kann es um Belohnung oder Vermeidung von Bestrafung gehen, zum jetzigen Zeitpunkt oder später. Jedoch lassen sich die funktionalen Anforderungen nicht vollständig auf hedonistische Ursachen zurückführen, sie berücksichtigen auch die Bedürfnisse der Gesellschaft, in der der Agent lebt.

In MAS werden funktionale Anforderungen vom Entwerfer vorgegeben, hedonistische Motive gibt es nicht. Die Funktionen sind hier rein zielorientiert, die Ziele sind vordefiniert oder ergeben sich aus der Funktion der Agenten im System.

Deontische Regeln sind mit Pflichten verbunden, d.h. mit Verboten und Idealen, die die Gesellschaft ihren Mitgliedern auferlegt. Sie haben oft die Form von Gesetzen oder Regulierungen, die festlegen, was getan werden soll und was nicht getan werden darf und damit gegebenenfalls die hedonistischen Motivationen der Agenten beschränken, indem sie Aktionen unterdrücken bevor sie ausgelöst werden.

5.6.4. Relationale Motivationen: Der Einfluss anderer Agenten

Der Ursprung relationaler Motivationen sind nicht der einzelne Agent selbst oder die Gesellschaft als Ganze, sondern andere Agenten. Eine relationale Motivation entsteht, wenn ein Agent einen anderen um etwas bittet oder etwas von ihm verlangt. Zwar kann der andere Agent die Bitte ablehnen, aber es gibt verschiedene Gründe, warum er dies nicht tut: Ablehnung ist nicht vorgesehen, Autoritätsbeziehungen, Gefallen u.a..

Relationale Motivationen haben eine Ähnlichkeit zu personalen Motivationen, denn einen anderen Agenten um etwas zu bitten kann analog dazu gesehen werden, sich selbst um etwas zu bitten. Der Unterschied liegt in dem dynamischen Charakter der Verbindungen zwischen den Agenten und insbesondere in der Beziehung zwischen dem bittenden und dem gebetenen Agenten, denn der letztere hat die Möglichkeit die Bitte abzulehnen, was ein Agent gegen sich selbst nicht kann.

Andererseits haben relationale Motivationen auch eine Ähnlichkeit zu deontischen Beziehungen. Die Bitte eine bestimmte Aufgabe auszuführen, etwa eine Dienstleistung, erzeugt eine innere Befriedigung und damit eine Tendenz zum Handeln. Die Aktion kann darin bestehen die Bitte zu erfüllen oder abzulehnen, in jedem Fall aber ist sie eine Antwort auf die Bitte, wie es deontische Regeln verlangen.

5.6.5. Verpflichtungen: Relationale und soziale Motivationen und Constraints

Wenn ein Agent eine Verpflichtung eingeht, bindet er sich selbst durch ein Versprechen oder einen Vertrag eine Aktion in der Zukunft auszuführen. Damit beschränkt er gleichzeitig eine eigenen Handlungsmöglichkeiten in der Zukunft. Die versprochene Aktion ist so zu sagen fest programmiert als ob das Verhalten der Agenten durch nichts mehr geändert werden könnte.

Alle menschlichen sozialen Systeme basieren auf relativ komplexen Verpflichtungsstrukturen. Verpflichtungen sind die Grundlage kollektiver Aktion. Es gibt sie in verschiedenen Formen und auf allen möglichen Ebenen der menschlichen Gesellschaft: Verträge, Konventionen, Eide, Versprechen, Verantwortung, Treue oder Pflicht, in der Familie, im Freundeskreis, im beruflichen Umfeld oder sogar im Staat. Sie erzeugen ein Netz feiner Abhängigkeiten und bauen dadurch eine soziale Organisation. Auf Grund von Verpflichtungen kann der Grad an Unvorhersagbarkeit für Aktionen anderer Agenten verringert werden, kann die Zukunft antizipiert werden und können die eigenen Aktionen geplant werden.

Formen von Verpflichtungen

Die verschiedenen Arten von Verpflichtungen werden nach den Analysedimensionen klassifiziert.

Relationale Verpflichtungen (ρ). Ein Agent verpflichtet sich gegenüber einem anderen, eine Aktion auszuführen oder für die Gültigkeit einer Information zu bürgen. Die Verpflichtung von *A* gegenüber *B* zu der Aktion *X* wird notiert durch

`commitDo(X,A,B)`

Die Zusicherung von *A* gegenüber *B*, dass die Aussage *P* gültig ist, wird notiert durch

`commitInfo(P,A,B)`

Sie besagt, dass *A* *P* annimmt und das sein Ziel ist, dass *B* ebenfalls *P* annimmt.

Umgebungsverpflichtungen (χ). Dies sind Verpflichtungen gegenüber Ressourcen. Ökologisches Bewusstsein ist ein Beispiel für eine Umgebungsverpflichtung.

Verpflichtungen gegenüber der sozialen Gruppe (σ). Hier gibt es zwei Untertypen. Der erste besteht in der Verpflichtung Aufgaben für die Gruppe auszuführen. Dieser Typ ist den relationalen Verpflichtungen ähnlich, nur dass die Gruppe als Ganze der Adressat ist, nicht einzelne Agenten. Verpflichtungen dieses Typs werden notiert durch

`commitDoGr(X,A,G)`

Der zweite Typ besteht in der Verpflichtung zu sozialen Konventionen und zur Übernahme der Constraints und der Aufgaben, die mit einer Rolle in einer Organisation verbunden sind. Ist R eine Rolle und G eine Gruppe, dann wird die Verpflichtung für die Rolle notiert durch

$$\texttt{commitRole(A,R)} =_{def} \forall \texttt{X} \in \texttt{tasks(R):}\ \texttt{commitDoGr(X,A,G)}$$

Verpflichtungen von Organisationen gegenüber ihren Mitgliedern (ϕ). Eine Organisation geht Verpflichtungen gegenüber ihren Mitgliedern ein, z.B. die Verpflichtung, sie zu bezahlen oder für ihren Schutz zu sorgen. Diese Art von Verpflichtungen stellen das Gegenstück zu den vorherigen dar. Sie werden notiert durch

$$\texttt{memberOf(A,G)} \Rightarrow \forall \texttt{X} \in \texttt{obligations(G): commitDoOrg(X,G,A)}$$

Verpflichtungen gegen sich selbst (π). Diese Verpflichtungen sind analog zu sehen zu den Verpflichtungen gegenüber anderen Agenten. Sie werden notiert durch

$$\texttt{commitDo(X,A,A)}$$

Verpflichtung als Strukturierung kollektiver Aktion

Verpflichtungen sind in der Regel nicht isoliert, sondern bilden eine Art von Abhängigkeitsstruktur, etwa in folgender Weise: *A* verpflichtet sich gegenüber *B* eine Aktion *X* auszuführen, woraufhin sich *B* gegenüber *C* verpflichtet eine Aktion *Y* auszuführen, wobei *B* weiß, dass *X* von *Y* abhängt.

Die Abhängigkeit der von *B* auszuführenden Aufgabe *Y* von der von *A* auszuführenden Aufgabe *X* wird notiert durch

$$\langle \mathrm{B}, \mathrm{Y} \rangle \leftarrow \langle \mathrm{A}, \mathrm{X} \rangle$$

Mit dieser Notation kann man Abhängigkeitsdiagramme oder Abhängigkeitsnetze über einer Menge von Abhängigkeiten bilden. Die Paare $\langle$`Agent,Aufgabe`$\rangle$ sind die Knoten und die Beziehungen zwischen ihnen die Kanten des Netzes.

Auch der Austausch von Informationen und Daten kann in Form von Abhängigkeitsbeziehungen beschrieben werden. Häufig dienen dann die Ausgabedaten des einen Prozesses als Eingabedaten für den nächsten. Falls ein Abhängigkeitsnetz Kreise enthält, kann es nicht ausgeführt werden. Es liegt dann ein Deadlock vor, weil eine Aufgabe von sich selber abhängt.

Verpflichtung als Antrieb für Aktion

Eine Verpflichtung ist ein Versprechen gegenüber einem anderen Agenten eine Aufgabe auszuführen, aber zugleich auch eine Selbstverpflichtung zur Ausführung der Aufgabe, d.h. man muss die Intention dazu haben. Dies gilt sowohl für Verpflichtungen gegenüber einzelnen Agenten als auch gegenüber der Gruppe. Die Abhängigkeit zwischen Verpflichtung und Intention wird in diesen beiden Fällen notiert durch

$$\texttt{commitDo(X,A,B)} \Rightarrow \texttt{commitDo(X,A,A)} \wedge \texttt{intention(A,X)}$$

$$\texttt{commitDoGr(X,A,B)} \Rightarrow \texttt{commitDo(X,A,A)} \wedge \texttt{intention(A,X)}$$

Ressourcenbeschränkungen und Kohärenz von Verpflichtungen

Verpflichtungen, die Ressourcen verbrauchen, schränken die Verfügbarkeit dieser Ressourcen für den Agenten selbst und für andere Agenten, zumindest zum selben Zeitpunkt, ein. Aus diesem Grund können Verpflichtungen widersprüchlich sein. Dies kann jedoch durch vorausschauende Planung vermieden werden. Es ist einer der Zwecke von Verpflichtungen die Agenten zur Planung von Aktionen zu veranlassen um so derartige Konflikte auszuräumen, bevor sie konkret werden. Es ist aber möglich eingegangene Verpflichtungen wieder rückgängig zu machen, wenn sich die Umstände, die zu einer Verpflichtung geführt haben, ändern oder wenn sich Prioritäten verschieben.

Gültigkeit und Persistenz von Verpflichtungen

Verpflichtungen bestehen so lange, bis eine der beiden folgenden Bedingungen erfüllt ist:

(1) Das Ziel der Verpflichtung ist erreicht worden.

(2) Der Agent bricht die Übereinkunft zu der gemachten Verpflichtung.

Der zweite Fall tritt ein, wenn der Agent seine Verpflichtung nicht mehr erfüllen kann, entweder weil er nicht mehr in der Lage dazu ist, oder weil sie von anderen Verpflichtungen abhängt, die die betreffenden Agenten nicht erfüllen. Grundsätzlich können Verpflichtungen gebrochen werden, deshalb gehört zu einer Verpflichtung auch die Verpflichtung andere Agenten zu informieren, wenn die Verpflichtung nicht eingehalten werden kann. Dies wird notiert durch

$$\texttt{commitDo(X,A,B)} \Rightarrow [\texttt{irgendwann}(\neg\texttt{canDo(A,X)}) \Rightarrow \texttt{inform(A,B,}\neg\texttt{canDo(A,X))}]$$

Bedeutung von Verpflichtungen

Verpflichtungen garantieren ein gewisses Maß an Stabilität für die Welt, indem sie den Agenten ein Gefühl der Sicherheit über das Verhalten der anderen Agenten oder auch der Organisation und eine feste Repräsentation der Welt geben. Auf diese Sicherheit hin können sie ihre eigenen Aktionen planen. Ohne Verpflichtungen könnten kognitive Agenten ebenso wenig in die Zukunft planen wie reaktive Agenten. Deshalb gehören Verpflichtungen untrennbar zum Konzept der kognitiven Agenten mit ihrer Fähigkeit zur Antizipation der Zukunft und mit ihrem Maß an freiem Willen.

5.7. Reaktive Ausführung von Aktionen

Im Entscheidungssystem eines Agenten wird festgelegt, welche potentielle Aktion tatsächlich ausgeführt wird. Aus der Menge der Tendenzen werden diejenigen ausgewählt, die von der Möglichkeit in die Handlung überführt werden sollen und es wird der Befehl zur Ausführung einer Aktion gegeben. Es sind verschiedene Arten von Entscheidungssystemen vorstellbar, die beiden wichtigsten sind:

- *Reaktive Ausführung von Aktionen*: Es wird nur der aktuelle Zustand berücksichtigt, d.h. die Information die im Augenblick vorliegt.
- *Intentionale Ausführung von Aktionen*: Der Agent projiziert sich selbst in die Zukunft und wertet die möglichen Konsequenzen von Aktionen aus, auch wenn dabei Ziele neu bewertet werden müssen.

In Abschnitt 5.7 wird die reaktive Ausführung behandelt, in Abschnitt 5.8 die intentionale Ausführung.

5.7.1. Konsumptive Aktionen und appetitives Verhalten

Eine konsumptive Aktion dient dazu eine Tendenz zu erfüllen, appetitives Verhalten ist die Suche nach einem Ziel, das es erlaubt, eine konsumptive Aktion auszuführen. Ein Beispiel ist die Nahrungsaufnahme. Die konsumptive Aktion ist hierbei das Essen, das appetitive Verhalten ist die Suche nach Nahrung. Beide gehören also zusammen, die konsumptive Aktion schließt das appetitive Verhalten ab, sie wird deshalb auch *finale Aktion* genannt. Der Unterschied zwischen beiden ist, dass das appetitive Verhalten auch bei Fehlen von Umweltreizen ausgeführt werden kann, während die konsumptive Aktion nur ausgeführt werden kann, wenn die entsprechenden Umweltreize vorhanden sind, diese dienen gewissermaßen als Vorbedingung für die Aktion.

Zwischen konsumptive Aktion und appetitives Verhalten kann noch ein „Taxi“ geschaltet sein. Damit sind Verhaltensweisen gemeint, die einen Agenten so orientieren und eventuell seinen Ort verändern, dass er die konsumptive Aktion ausführen kann. Ein Taxi dient also dazu ein Ziel zu

erreichen. Typischerweise benutzen Taxis Gradienten-Verfolgungs-Mechanismen um das Ziel zu erreichen. Die Abfolge von appetitivem Verhalten, Taxi-Verhalten und konsumptiver Aktion nennt man eine Verhaltenskette.

5.7.2. Aktionsauswahl und Steuerungsmodi

Die Probleme, die bei der Auswahl einer Aktion durch einen reaktiven Agenten bestehen, lassen sich in drei Punkten zusammenfassen:

1. Was soll der Agent tun, wenn mehrere Verhaltensketten miteinander konkurrieren, z.B. wenn sich der Agent zwischen Essen, Trinken oder Paaren entscheiden muss?
2. Was soll der Agent tun, wenn er mehrere verschiedene Triebe in sich fühlt und eine ganze Menge von Stimuli bekommt, z.B. wenn er gerade auf Futtersuche ist und ein attraktives anderes Wesen seiner Gattung sieht, mit dem er sich paaren könnte?
3. Wie soll der Agent auf neue Umgebungsbedingungen achten, die für ihn möglicherweise lebensbedrohend sind oder die ihn daran hindern sein Ziel zu erreichen, wie z.B. das Vorhandensein eines Jägers oder eines Hindernisses?

Ein gutes Auswahlmodul sollte folgende Kriterien erfüllen (Tyrrell und Werner):

1. Ausreichende Allgemeinheit um ad-hoc-Modelle zu vermeiden.
2. Eine gewisse Persistenz der konsumptiven Aktionen um Unbeständigkeit des Agenten zu vermeiden. Dies wird dadurch erreicht, dass man den externen Stimuli eine ausreichende Intensität gibt, abhängig von der Distanz zwischen dem Agenten und dem gewünschten Objekt.
3. Priorisierung auf Grund von Motivationen.
4. Bevorzugung konsumptiver vor appetitiven Aktionen. Hat der Agent etwas, dann braucht er nicht woanders danach zu suchen.
5. Bevorzugung vollständig durchgeführter Aktionsfolgen vor immer neu begonnenen aber nicht vollendeten Aktionsfolgen.
6. Unterbrechung einer Aktion wenn eine neue Situation den Agenten bedroht (Jäger oder andere Gefahr) oder wenn die Aktion offensichtlich nicht vollständig durchgeführt werden kann (Hindernis).
7. Opportunismus, Ausführung konsumptiver Aktionen wenn möglich, auch wenn sie nicht Bestandteil der aktuellen Aktionsfolge sind.
8. Ausnützung aller verfügbaren Informationen um das in einer Situation günstigste Verhalten zu bestimmen.
9. Parallelausführung von Aktionen, z.B. Flucht und Aussenden eines Warnsignals.
10. Einfache Erweiterungsmöglichkeit zwecks Hinzufügen neuer Aktionen und Aktionsfolgen.
11. Lernfähigkeit, z.B. durch geeignete Modifikation von Verhaltensoptionen auf Grund von Erfahrungen.

Situierte Aktionen und Potentialfelder

Diese beiden Modelle wurden in Kapitel 4 behandelt. Das Modell der situierten Aktionen benutzt eine Menge von Regeln, in denen Aktionen mit Situationen verknüpft werden. Die Situationen definieren vor allem Umgebungsbedingungen, aber man kann auch Parameter einführen, die internen Zuständen entsprechen, und somit die Situationen als Paare der Form

⟨*Umgebungsbedingungen*, *interner Zustand*⟩

definieren. Situierte Aktionen haben aber einige Nachteile. Erstens ist es schwierig mit ihnen echte Ausdauer darzustellen, zweitens können sie nicht gleichzeitig sehr opportunistisch sein und zu Entscheidungen auf der Basis von Gewichtungen fähig und drittens ist es damit schwierig einen Agenten zu implementieren, der mehrere verschiedene Aufgaben durchführen kann.

Potentialfelder haben einige Vorzüge. Da die Distanz zwischen dem Agenten und dem Ursprung der Stimuli berücksichtigt wird, statten die Felder einen Agenten mit einer gewissen Persistenz aus. Sie erlauben es ferner neue Informationen zu berücksichtigen, opportunistisch zu sein und mehrere Aktionen parallel durchzuführen. Um bei Potentialfeldern interne Motivationen zu berücksichtigen braucht man nur die Intensität eines Signals mit einem Faktor zu multiplizieren, der von den Motivationen abhängt.

Die Nachteile der Potentialfelder sind, dass sie nicht alle Probleme bei der Auswahl von Aktionsfolgen lösen können, da sie sehr stark auf die Umgebung bezogen sind, und dass es manchmal nötig ist, andere Mechanismen zu verwenden um ein effektives Verhalten zu erzielen.

Hierarchische Modelle

In diesen Modellen werden Verhaltensweisen auf höheren Ebenen in Aktionen auf niedrigeren Ebenen zerlegt. Das berühmteste Modell dieser Art ist das des Ethologen Tinbergen (1951). In ihm ist das Verhalten in Form einer Hierarchie von Knoten organisiert. Die Knoten kann man sich wie formale Neuronen vorstellen. Mit jedem Knoten ist ein *angeborener Triggermechanismus* (innate trigger mechanism, ITM) verbunden, der eine Art Vorbedingung für die Aktivierung der Knoten darstellt. Ein Knoten werden durch eine Kombination von Stimuli aktiviert, die von außen und von den hierarchisch übergeordneten Knoten kommen. Wenn die Zahl der eingehenden Stimuli einen Schwellenwert überschreitet, wird der Knoten „ausgelöst“ und seine Aktivierungsenergie wird an die Knoten auf einer tiefer liegenden Ebene übertragen. Höhere Knoten entsprechen allgemeineren Aktivitäten wie *Reproduzieren* oder *Reinigen*, und je tiefer ein Knoten in der Hierarchie liegt, desto detaillierter ist die ihm entsprechende Aktion, z.B. Entspannen der Muskel. Die Knoten auf der selben Ebene stehen in Konkurrenz zueinander, d.h. wenn einer aktiviert ist, werden die anderen gehemmt. Abbildung 5.6 zeigt als Beispiel die hierarchische Gliederung der Instinkte nach Tinbergen.

Rosenblatt/Payton und Tyrrell haben andere Modelle vorgeschlagen, die zwar auch eine hierarchische Struktur besitzen, aber bei der Entscheidungsfindung nicht hierarchisch arbeiten. Die Hierarchien dieser Modelle werden *Freiflusshierarchien* genannt, weil die übergeordneten Knoten nicht entscheiden, welche der untergeordneten Knoten aktiviert werden, sie definieren nur Präferenzen in einer Menge von Kandidaten auf niedrigerer Ebene. Abbildung 5.7 zeigt eine solche Hierarchie für die Futtersuche.

Bottom-up-Modelle

Bei den Bottom-up-Modellen wird bewusst jede Form von Dominanz ausgeschlossen. Ein Vertreter dieser Modelle ist das ANA-System von Pat Maes. Es besteht aus einem Netz von Modulen, wobei jedes Modul eine elementare Aktion repräsentiert, wie `gehe-zum-Futter`, `trink-Wasser`, `flüchte-vor-Kreatur` usw. Jede Aktion wird durch fünf Angaben definiert:

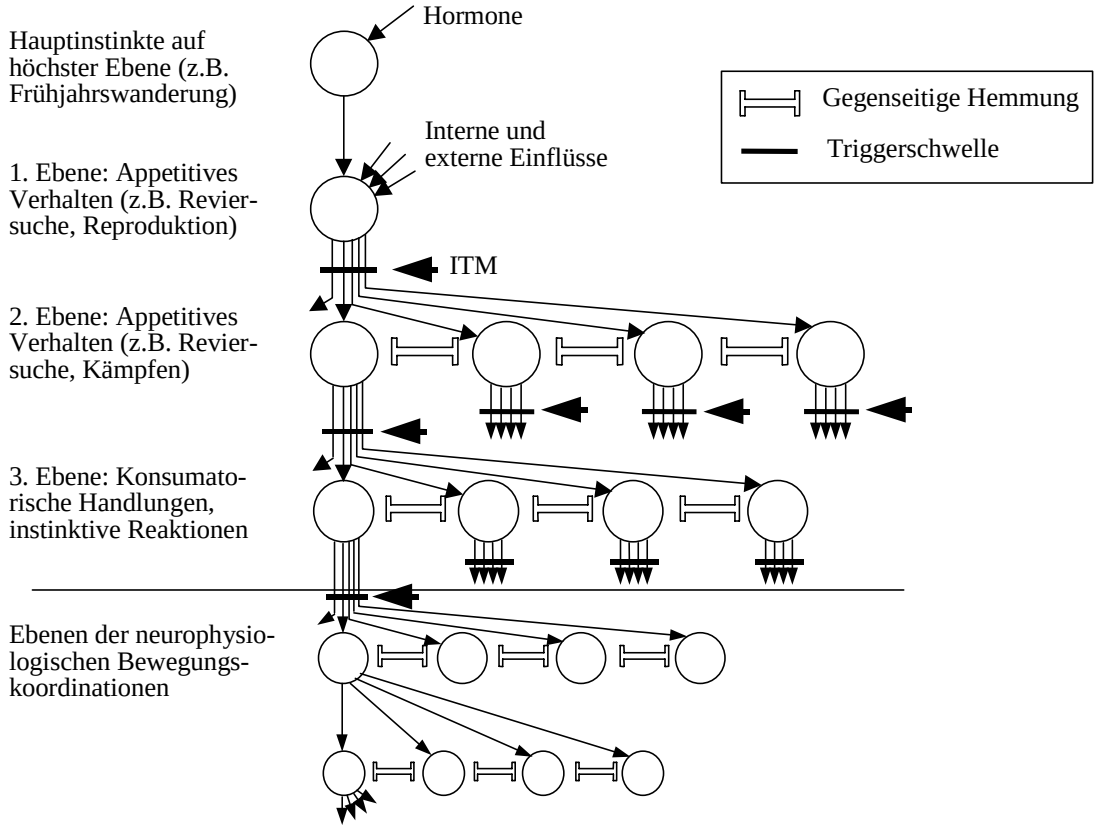

Abbildung 5.6

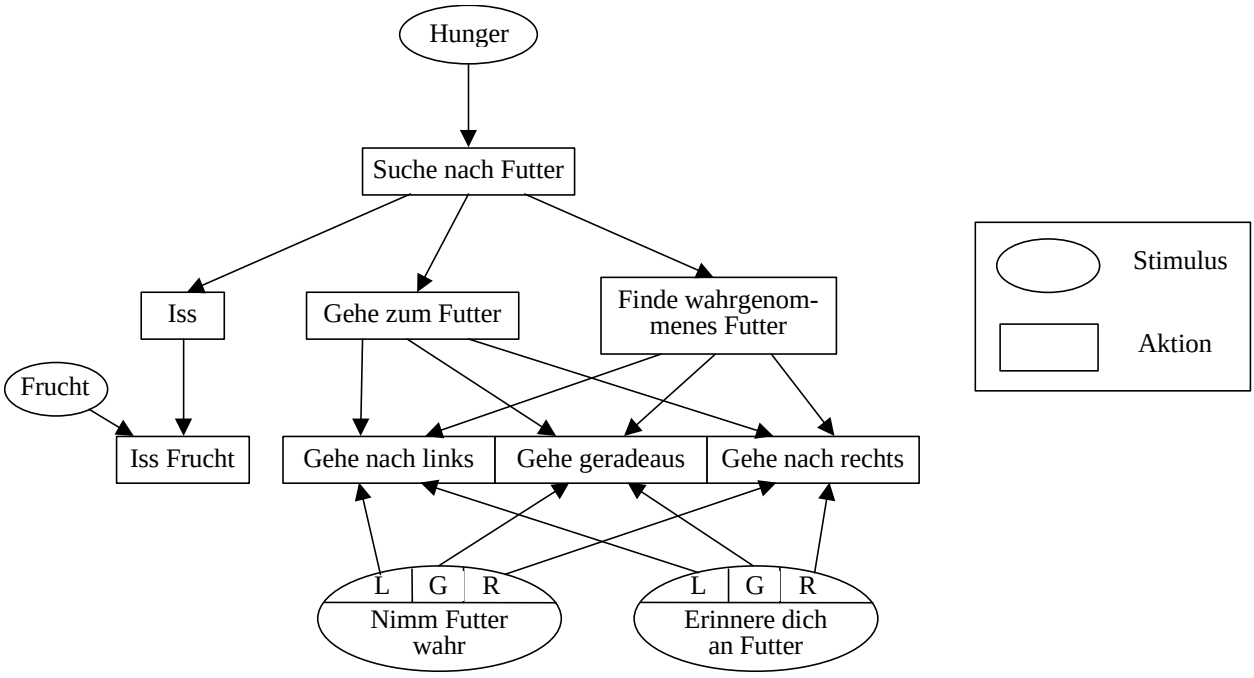

Abbildung 5.7

- ihre Anwendungsbedingungen
- ihr Aktivierungsniveau
- ihre Kanten zu den Vorgängern
- ihre Nachfolger
- Aktionen, zu denen sie in Gegensatz steht

Abbildung 5.8 zeigt ein Netz, das die Aktion `gehe-zu-Kreatur` repräsentiert. Die Stimuli der Umgebung und die internen Motivationen (Furcht, Müdigkeit, Aggression, Hunger, Durst, Neugier, Faulheit) tendieren dazu bestimmte Aktionen zu aktivieren, indem sie ihr Aktivierungsniveau anheben. Die Aktion mit dem höchsten Niveau wird ausgewählt und das Aktivierungsniveau ihrer Nachfolger wird etwas angehoben, wodurch ihre Chance ausgewählt zu werden steigt.

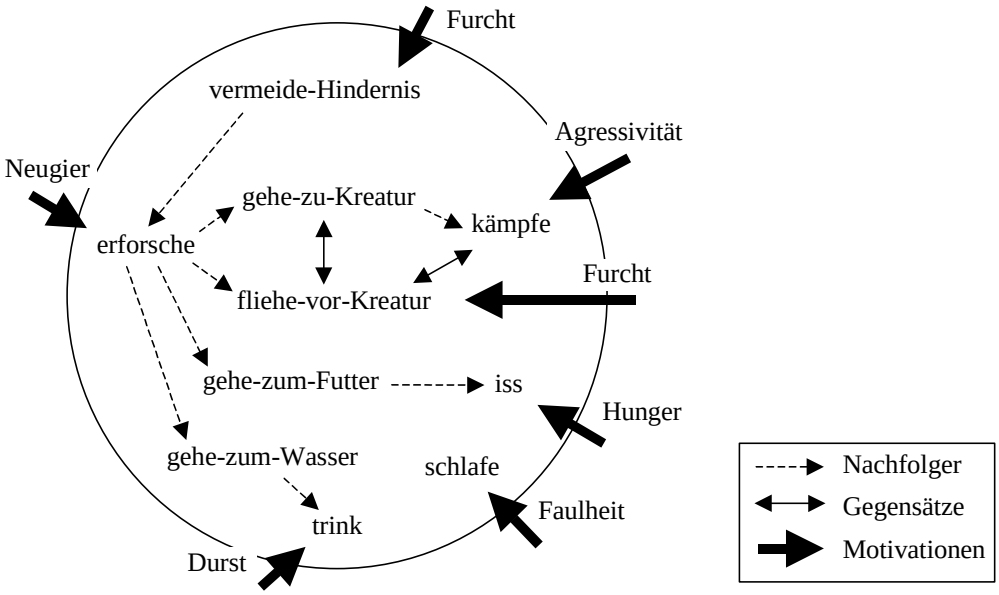

Abbildung 5.8

Werners Modell ist an Neuronalen Netzen orientiert. Interne und externe Motivationen werden multiplikativ kombiniert. Das Modell enthält vier Mengen von Knoten, vgl. Abbildung 5.9:

- Eingabeknoten, die mit Typen von Perzepten verknüpft sind, die auf den Agenten (`Futter-rechts`, `Jäger-voraus`) und auf Triebe bezogen sind (Hunger, Durst usw.),
- Ausgabeknoten für elementare Aktionen (gehe-nach-rechts, gehe-nach-links),
- Knoten, die verborgenen Knoten im Netz entsprechen,
- Knoten, die Befehlen entsprechen. Die Befehle werden als Koeffizienten benutzt, die eine multiplikative Verknüpfung zwischen den Eingabe- und Ausgabeknoten bewirken.

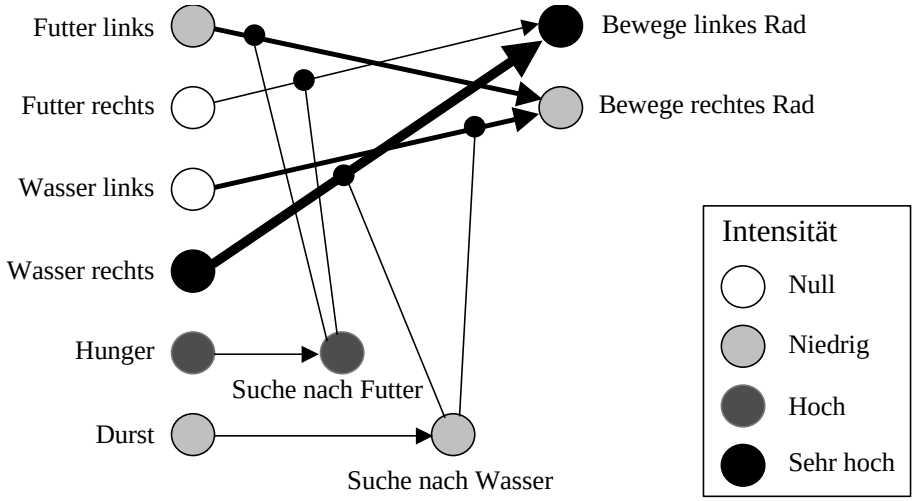

Abbildung 5.9

Drogouls EMF-Modell (Etho Modelling Framework) unterscheidet sich von den anderen Ansätzen dadurch, dass es nicht einen einzelnen Agenten isoliert betrachtet, sondern in Interaktion mit anderen Agenten. Die Betonung liegt hier auf der Fähigkeit zu sozialer Anpassung und zur Spezialisierung der Fähigkeiten des Agenten im zeitlichen Verlauf. Es wird nicht zwischen appetitiven Aktionen, Taxis und konsumptiven Aktionen unterschieden, vielmehr zwischen Aufgaben und Aktionen. Eine Aktion ist eine elementare Verhaltensweise, z.B. einem Gradienten folgen, ein Objekt ergreifen oder Futter zu sich nehmen. Eine Aufgabe ist eine Menge von Aktionen, die in einer bestimmten Reihenfolge ausgeführt werden, wie in einem klassischen Programm. Die Ausführung einer Aufgabe kann zu jedem Zeitpunkt gestoppt werden, wenn Umgebungsbedingungen oder der Zustand des Agenten dies erforderlich machen.

Eine Aufgabe wird durch eine Kombination interner und externer Stimuli ausgelöst. Sie besitzt ein Gewicht, das bei jeder Ausführung der Aufgabe erhöht wird, wodurch eine Tendenz zur Spezialisierung auf bestimmte Aufgaben hervorgerufen wird. Abbildung 5.10 zeigt ein Diagramm des Steuerungsmechanismus für Aufgaben. Interne und externe Stimuli werden multipliziert, außerdem wird das Gewicht der Aufgabe als Multiplikator benutzt um ihren Einfluss entsprechend ihrem Gewicht geltend zu machen. Der Komparator vergleicht alle Aufgaben, deren Gewicht über einem bestimmten Schwellenwert liegt. Das Gewicht der ausgewählten Aufgabe wird um einen bestimmten Betrag erhöht.

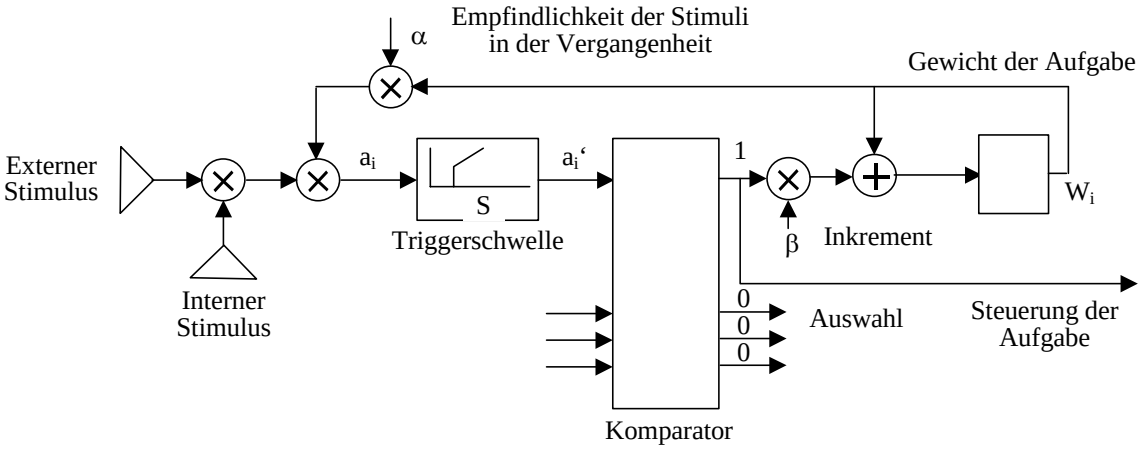

Abbildung 5.10

5.7.3. Aktionsauswahl oder dynamische Kombination

Um die Beschränkungen der kompetitiven Aktionsauswahl zu vermeiden, definierte L. Steels sein dynamisches Systemmodell, vgl. Kapitel 3. Das Modell kann durch das Diagramm von Abbildung 5.11 illustriert werden.

Die elementaren Verhaltensweisen sind (meist lineare) Funktionen interner und externer Stimuli und der Werte der Realisierer (Geschwindigkeit der Räder, Zustand des Greifers usw.). Die Ergebnisse der Funktionen werden addiert und den Realisierern übergeben. Die formale Definition eines Realisiererwertes ist durch die folgende dynamische Gleichung gegeben:

$$y_i(t+1) = \sum_{k=1}^{p} f_k\left(x_1(t),\ldots,x_n(t),y_1(t),\ldots,y_m(t)\right)$$

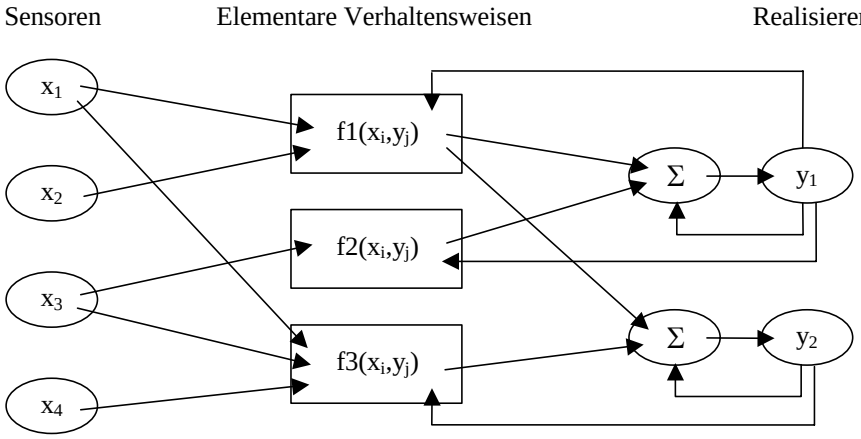

Abbildung 5.11

In manchen Fällen kann man die Technik der Kombination von Aktionen unabhängig von einer konkreten Architektur verwenden. Das ist dann der Fall, wenn sich Aktionen als Vektoren in einem metrischen Raum darstellen lassen. Die Kombination von Aktionen lässt sich dann als Vektoraddition darstellen.

Diese Technik wurde von Zeghal in einem System verwendet, das den Zweck hatte, Kollisionen zwischen Flugzeugen zu vermeiden. Jedes Flugzeug betrachtet das andere als eine potentielle Gefahr. Wenn die Gefahr eines Zusammenstoßes besteht, versuchen die Flugzeuge diesen zu vermeiden, indem sie einander umgehen und dabei trotzdem ihre Flugroute einhalten. Wenn sie sich aber zu nahe kommen, können sie voreinander fliehen, auch wenn sie dazu ihre Richtung ändern müssen. Das Verhalten der Flugzeuge kann durch den folgenden Vektor beschrieben werden:

$$\vec{X} = \frac{\alpha\vec{B} + \beta\vec{F} + \delta\vec{C}}{\alpha + \beta + \delta}$$

Die Vektoren $\vec{B}$, $\vec{F}$ und $\vec{C}$ repräsentieren die Anziehung durch das Ziel, Flucht und Umgehen. Die Stärke der Vektoren hängt von dem Potential ab, das durch die Ziele und die Gefahren erzeugt wird. Die Parameter α, β und δ entsprechen den kritischen Faktoren, die von den jeweiligen Positionen des Agenten, des Ziels und vor allem der Gefahrenstelle abhängen. Z.B. nimmt β mit abnehmender Distanz zwischen Gefahrenstelle und Agent zu, wobei sich α und δ abschwächen. Damit ist es möglich neue Verhaltensweisen einfach hinzuzufügen, z.B. das Fliegen in Geschwadern, in denen die Agenten Gruppen bilden, und das gruppenbildende Verhalten wird höher gewichtet als die übrigen Verhaltensweisen.

5.8. Intentionale Ausführung von Aktionen

Intentionen können nicht auf einfache Konflikte zwischen Tendenzen, entstanden aus Trieben, Wünschen und Beschränkungen, reduziert werden. Tendenzen haben ihren Ursprung in den sehr einfachen Elementen der Psyche eines Agenten, Intentionen dagegen setzen eine überlegte Analyse der gegenwärtigen und zukünftigen Situation und der Möglichkeiten für eine Aktion voraus. Kognitive Agenten, auch intentionale Agenten genannt, können bis zu einem gewissen Grad die Konsequenzen ihrer Entscheidungen voraus sagen, weil sie sich kausale Beziehungen zwischen Ereignissen vorstellen können und damit die Zukunft vorweg nehmen können. Ein intentionaler Agent ist also notwendigerweise kognitiv, denn er trifft seine Entscheidungen durch Schlussfolgern über einem Weltmodell. Die Entscheidungsfindung ist auch rational insofern, als sie den Zielen des Agenten unterworfen ist. Durch bewusste Berechnung kann der Agent verschiedene Situationen

vergleichen, die er voraus berechnen kann, und er kann zwischen den möglichen Ergebnissen vergleichen und daraufhin entscheiden, welche Situation am besten seinen Zielen entspricht.

5.8.1. Logische Theorien der Intention

Man muss zwischen zwei Arten von Intention unterscheiden, der Intention in der Zukunft zu handeln und der Intention etwas jetzt zu tun. Die erste Art ist auf die Zukunft gerichtet und ist die Quelle für ein Verhalten, insbesondere für andere Intentionen, die sich aus der ursprünglichen ergeben um das gewünschte Ziel zu erreichen. Die zweite Art bezieht sich auf die unmittelbare Aktion, die Aktion muss sofort ausgeführt werden und impliziert keine weiteren Intentionen.

Die beiden Arten der Intention haben zwar Ähnlichkeiten miteinander, sie fallen aber nicht zusammen. Die erste setzt einen Planungsmechanismus voraus, die zweite bezieht sich nur auf den Mechanismus der Aktionsausführung. Die erste Art ist interessanter im Hinblick auf intentionale Agenten, deshalb wird nur sie betrachtet. Intention wird also im Folgenden als der Mechanismus der Projektion in einen zukünftigen Weltzustand, der noch nicht realisiert ist, verstanden.

Die wesentlichen Begriffe zur Beschreibung von Intentionen sind *Zeit*, *Aktion* und *Annahmen*. Ein Agent will eine Aktion ausführen, weil er annimmt, dass er dadurch ein Ziel erreichen kann. Es werden deshalb Definitionen für Aktion, Ereignis und Weltzustand benötigt und dafür wie Weltzustände und Ereignisse zeitlich miteinander verknüpft sind. Bratman gibt einige Merkmale an, die für eine Theorie der intentionalen Aktion notwendig sind. Danach sind Intentionen der Ursprung von Problemen, die Agenten lösen müssen. Intentionen sollten ferner kohärent untereinander sein, was impliziert, dass sie die Annahme anderer Intentionen erlauben müssen. Agenten müssen ihre Aktionen bei der Ausführung einer Intention verfolgen können um im Bedarfsfall zur Erfüllung der Intention umplanen zu können. Schließlich muss noch bei der Ausführung zwischen den gewollten Ergebnissen und den nur zufällig als Nebeneffekt entstandenen unterschieden werden können.

Definition *Intention*
Ein Agent `x` hat die Intention eine Aktion `a` auszuführen, geschrieben `intention(x,a)`, wenn `x` das Ziel hat, dass eine Aussage `P` über den Zustand der Welt wahr werden soll, geschrieben `goal(x,P)`, und dass die folgenden Bedingungen erfüllt sein sollen: (1) `x` nimmt an, dass `P` eine der Konsequenzen von `a` ist; (2) `x` nimmt an, dass `P` zur Zeit nicht gilt; (3) `x` nimmt an, dass er `a` ausführen kann; (4) `x` nimmt an, dass `a` möglich ist und dass somit `P` erfüllt werden kann.

Das Problem dieser Definition ist, dass sie von dem Konzept abhängt zu ‚glauben, dass eine Aktion möglich ist'. In einer statischen Welt mit nur einem Agenten ist eine Aktion ausführbar oder möglich, wenn die Bedingungen für ihre Ausführung gegenwärtig realisierbar sind oder wenn ein Plan existiert, d.h. eine Folge von Aktionen, die von dem Agenten ausgeführt werden können, der es möglich macht, die Bedingungen für die Aktion zu erfüllen, die der Agent auszuführen beabsichtigt. Ist die Welt aber nicht statisch, d.h. entwickelt sie sich, oder gibt es mehrere Agenten, die die Welt durch ihre jeweils eigenen Aktionen verändern, und es ist nicht sicher, dass der Agent seine Aktion vollständig ausführen kann, dann sind die Bedingungen für die Ausführung einer Aktion weniger einfach, denn sie hängen davon ab, dass man die Entwicklung der Welt antizipieren kann.

In einer Welt mit mehreren Agenten muss ein Agent außerdem noch über die Fähigkeiten und Intentionen der anderen Agenten Bescheid wissen um den Zustand der Welt und damit die Bedingungen für eine eigene Aktion antizipieren zu können.

Aus diesen Gründen können in einer sich entwickelnden Welt mit mehreren Agenten Intentionen nur unter der Annahme regelmäßiger Abläufe entwickelt werden, d.h. auf Grund der Tatsache, dass die Umgebung und die anderen Agenten sich an die Aufgaben, die sie in der Zukunft auszuführen beabsichtigen, gebunden fühlen. Jede Definition von Intention in einem kognitiven Multiagentensystem beruht auf der Möglichkeit gemeinsame Aktionen mehrerer Agenten in einer sich entwic??kelnden Welt zu planen. Außerdem wird noch vorausgesetzt, dass sich ein Agent zur Ausführung einer Aktion verpflichtet fühlt, anderen oder sich selbst gegenüber. Deshalb ist das Konzept der Verpflichtung (commitment) ein grundlegendes Element bei der Beschreibung des Übergangs eines kognitiven Agenten zu einer Aktion.

5.8.2. Cohens und Levesques Theorie der rationalen Aktion

Notationen

Der Formalismus von Cohen und Levesque basiert auf der Modallogik. Diese ist die Logik erster Ordnung (mit Gleichheit), erweitert um einige Operatoren, die propositionale Haltungen oder Folgen von Ereignissen repräsentieren. Die beiden grundlegenden Modalitäten sind Annahme und Ziel:

- `believe(x,p)` - Agent `x` annimmt, dass `p` gilt,
- `goal(x,p)` - Agent hat `x` das Ziel `p` zu erreichen.

Es gibt noch weitere Prädikate, die sich auf das Konzept *Ereignis* beziehen. Aktionen werden generell als Ereignisfolgen betrachtet, die die Erfüllung von Aussagen ermöglichen. Die Prädikate berücksichtigen keine Seiteneffekte oder Konsequenzen von Aktionen oder das Problem der Gleichzeitigkeit. Es sind folgende Prädikate:

- `agent(x,e)` - Agent `x` ist der einzige Agent, der die Ereignisfolge `e` initiiert,
- `e1` $\leq$ `e2` - `e1` ist eine Ereignisfolge, die `e2` vorausgeht,
- `happens(e)` - die Ereignisfolge `e` geschieht gerade,
- `done(e)` - die Ereignisfolge `e` ist gerade beendet worden.

Auf den Ereignisfolgen oder Aktionen sind spezielle Operatoren definiert, die ähnlich zu denen in der dynamischen Logik sind:

- `a;b` - sequentielle Komposition von Aktionen. Das heißt, zuerst wird `a` ausgeführt und erst wenn es vollständig ausgeführt ist, wird `b` ausgeführt.
- `a|b` - nichtdeterministische Auswahl von Aktionen. Das heißt, es wird entweder `a` oder `b` ausgeführt.
- `p?` - Aktion, die testet, ob `p` erfüllt ist. Ist `p` nicht erfüllt, dann blockiert die Aktion andere Aktionen.
- `a*` - Wiederholung. Aktion `a` wird so lange wiederholt, bis sie durch eine nicht erfüllte Eigenschaft unterbrochen wird.

Mit Hilfe der elementaren Definitionen können komplexere Kontrollstrukturen definiert werden:

- `if p then a else b` := `p?;a|¬p?;b`
  Das Konditional ist eine Aktion, bei der `a` ausgeführt wird so bald `p` erfüllt ist, und `b`, so bald `p` nicht erfüllt ist.
- `while p do a` := `(p?;a)*¬p?`
  Aktion `a` wird so lange wiederholt, wie `p` wahr ist und anschließend ist `p` falsch.
- `eventually(p)` := `∃e happens(e;p?)`
  Der Operator *eventually* beschreibt, dass es eine Aktion `e` gibt, nach deren Ausführung `p` wahr sein wird.
- `always(p)` := `¬eventually(¬p)`
  Der Operator *always* beschreibt, dass `p` immer wahr ist.
- `later(p)` := `¬p ∧ eventually(p)`
  `p` ist im Augenblick nicht wahr, wird aber später wahr werden.
- `before(p,q)` := `∀c happens(c;q?) ⇒ ∃a (a≤c) ∧ happens(a;p?)`
  Wenn `q` wahr ist, ist `p` schon vorher wahr gewesen.

Als Abkürzungen werden verwendet:

```
done(x,a) := completed(a) ∧ agent(x,a)
happens(x,a) := happens(a) ∧ agent(x,a)
```

Alle diese Operatoren werden axiomatisch definiert (Axiomensystem S5) und ihre Semantik wird mittels möglicher Welten festgelegt. Weitere Aussagen lassen sich als logische Sätze auf der Basis der Definitionen formulieren, z.B.

```
eventually(q) ∧ before(p,q) ⇒ eventually(p)
```

Intention und rationale Aktion

Intention wird als Auswahl (einer Aktion) und Verpflichtung zu dieser Aktion verstanden. *Ziel* wird nicht im üblichen Wortsinn verstanden, sondern formal dadurch definiert, dass `goal(x,p)` bedeutet, dass `p` in allen möglichen Welten wahr ist, in denen ein bestimmtes Ziel erreicht werden kann. Das heißt praktisch, dass statt einer Definition des Begriffs *Ziel* auf die Konsequenzen von Zielen rekurriert wird. Intention kann auch als ein persistentes Ziel betrachtet werden, d.h. eines, das ein Agent so lange beibehält wie bestimmte Bedingungen bestehen. Dieser Aspekt wird in folgender Weise definiert:

```
p-goal(x,p) := goal(x,later(p)) ∧ believe(x,¬p) ∧
               before( believe(x,p) ∨ believe(x,always(¬p)),
                       ¬goal(x,later(p))
```

Persistente Ziele schließen also eine Verpflichtung ein, wenn auch nur indirekt. Ein Agent hält an seinem Ziel fest so lange nicht eine bestimmte Bedingung erfüllt ist. Man kann also annehmen, dass ein Agent nicht ewig an einem Ziel festhält, deshalb kann man aus der Definition des persistenten Ziels die folgende Aussage ableiten:

```
p-goal(x,p) ⇒ eventually[believe(x,p) ∨ believe(x,always(¬p))]
```

Es gibt drei Arten von Intentionen. Die ersten beiden beschreiben die Absicht eines Agenten x eine Aktion a auszuführen. Die erste Art ist die „fanatische“ Verpflichtung a auszuführen so bald dies möglich ist, genauer gesagt, so bald der Agent glaubt, sie sei ausführbar.

```
fan-intention(x,a) :=  p-goal(x,done(x,believe(x,happens(a))?;a))
```

Die zweite Art von Intention wird relativ zu einer Bedingung definiert, die erfüllt sein muss, damit ein Agent die Intention hat. Für ihre Definition wird das Konzept des *relativ persistenten Ziels* benötigt. In ihm werden gewissermaßen Gründe für einen Agenten, eine Aktion auszuführen, festgelegt.

```
p-r-goal(x,p,q) :=
     goal(x,later(p)) ∧ believe(x,¬p) ∧
     before( believe(x,p) ∨ believe(x,always(¬p)) ∨ believe(x,¬q),
            ¬goal(x,later(p))

rel-intention(x,a,q) :=
     p-r-goal(x,done(x,believe(x,happens(a))?;a),q)
```

Die dritte Art von Intention ist die *vage* oder *abstrakte Intention*. Intentionen dieser Art beziehen sich auf die Erfüllung einer Aussage p und nicht auf die Ausführung einer Aktion. Ein Agent, der eine Intention dieser Art hat, nimmt an, dass es eine Folge von Aktionen e‘ gibt, die schließlich die Ausführung der Aktion e ermöglicht, die p erfüllt.

```
abs-intention(x,p) :=
     p-goal(x,∃e done(x,[ believe(x,∃e‘ happens(x,e‘;p?) ∧
                          ¬goal(x,¬happens(x,e;p?))]?;e;p?))
```

Diskussion

Cohens und Levesques Theorie hat den Vorteil einer strikten Formalisierung rationaler Aktionen mit Hilfe einer relativ einfachen und präzisen Sprache. Viele Situationen können damit beschrieben werden. Es gibt einige weiterentwickelte Theorien auf dieser Basis, z.B. die Theorie der Kooperationsaktionen von Galliers oder die Theorie sozialer Intentionen und Einflüsse von Castelfranchi und Conte. Die Schwächen der Theorie liegen in folgenden Punkten:

(1) Aktionen werden als Ereignisse oder Ereignisfolgen betrachtet, nicht als auf die Welt anwendbare Operationen. Die von den Aktionen der Agenten bewirkten Modifikationen werden daher nicht berücksichtigt, deshalb ist die Theorie zur Lösung von Planungsproblemen und zur Behandlung von Phänomenen, die sich auf die Interaktion zwischen Agenten beziehen, nicht geeignet. Mit einer Erweiterung der Theorie um Kommunikationsoperatoren lassen sich aber Interaktionen zwischen Agenten behandeln.

(2) Mechanisches Beweisen von Aussagen (Theorembeweisen) ist in Theorien dieses Typs schwierig. Beweise müssen praktisch „von Hand“ gemacht werden mittels semantischer Betrachtungen unter Einbeziehung möglicher Welten.

(3) Die Theorie liefert keine operative Beschreibung der Mechanismen, die bei der Ausführung von Aktionen durch rationale Agenten eine Rolle spielen, vielmehr nur die minimalen Bedingungen für die Intention eines Agenten.

Die Theorie beschreibt einige wesentliche Konzepte von Intention. Sie hat aber die Tendenz im Hinblick auf Aktionen sehr einfach zu sein, sie berücksichtigt z.B. nicht die Revision von Aktionen, sie besitzt auch keinen Mechanismus zur Entwicklung der Intentionen und der Konflikte, die durch verschiedene Motivationen und Constraints entstehen können. Sie nimmt an, dass eine 1:1-Beziehung zwischen Zielen und Aktionen zum Erreichen der Ziele besteht und kann den häufigen Fall, dass ein Ziel auf verschiedenen Wegen erreicht werden kann, nicht richtig erfassen. Sie sagt nichts aus über Entscheidungsfindung in dem Fall, dass verschiedene Wege bestehen. Man kann nur ein Ziel aufgeben, wenn es nicht mehr geeignet erscheint, aber dafür muss man explizit Gründe angeben. Eine gute Theorie rationaler Aktion sollte die Konzepte von Zielen, Annahmen, Entscheidungen, Ausführung von Aktionen, Verpflichtungen, Kommunikationen und Ergebnissen von Aktionen integrieren und koordinieren.

Shohams Ansatz des „agentenorientierten Programmierens“ entspricht diesem Ziel. Er definiert eine agentenorientierte Programmiersprache als eine Art Erweiterung der objektorientierten Sprachen. Die Erweiterung besteht darin, dass zur Programmierung der Agenten Konzepte wie Ziele, Entscheidungen, Fähigkeiten, Annahmen usw. verwendet werden. Die Typen der Nachrichten, die die Agenten austauschen, definieren Nachrichten über Daten, Anforderungen, Angebote, Versprechen, Ablehnungen, Annahmen usw. und beziehen sich somit auf Kommunikationsmechanismen auf hoher Ebene.

Der Programmiersprache liegt eine modallogische Sprache mit einer Semantik möglicher Welten zu Grunde. Die beiden grundlegenden Modalitäten sind

$believe(x,p)^t$	-	Agent x glaubt, dass die Aussage p zur Zeit t wahr ist
$commitment(x,y,a)^t$	-	Agent x hat sich zur Zeit t dem Agenten y gegenüber zur Ausführung der Aktion a verpflichtet

Alle anderen Modalitäten werden von diesen beiden abgeleitet. Zum Beispiel ist die Entscheidung zur Ausführung einer Aktion als Verpflichtung gegen sich selbst definiert und die Fähigkeit zur Ausführung als die Implikation, dass die Auswahl von p zu p führt:

$$choice(x,p)^t := commitment(x,x,p)^t$$
$$capable(x,p)^t := choice(x,p)^t \Rightarrow p$$

6. Kommunikation

6.1. Aspekte der Kommunikation

6.1.1. Zeichen, Indizes und Signale

Kommunikation erfolgt mittels Signalen; jedes *Signal* charakterisiert einen *Index* oder ein *Zeichen*. Diese drei Begriffe müssen genau unterschieden werden. Das Signal ist das einfachste Element in der Kommunikation. Es kann als Marke oder als Spur oder als Modifikation der Umgebung auf der mechanischen Ebene (hören, fühlen), der elektromagnetischen Ebene (sehen) oder der chemischen Ebene (schmecken, riechen) definiert werden. Es kann Information für diejenigen transportieren, die es empfangen können. Signale breiten sich nach den jeweiligen physikalischen Gesetzen aus und haben für sich genommen keine Bedeutung. Damit sie Bedeutung erhalten, müssen zwei Bedingungen erfüllt sein: Es muss ein Agent existieren, dessen Wahrnehmungsfähigkeit zum Empfang der Signale geeignet ist, und ein Interpretationssystem, das das Signal in Bedeutung umwandeln kann oder zumindest in ein Verhalten.

Wird das Signal in ein Verhalten umgewandelt, dann nennt man es einen *Stimulus*. Das ist bei den meisten Tieren der Fall sowie bei reaktiven Agenten. Ein solches Signal kann eine Tendenz in einem Agenten verstärken. Ist diese ausreichend stark, dann führt das Signal zu einer Handlung. Erzeugt das Signal kein Verhalten, sondern wird in das kognitive System eingefügt, dann nennt man es einen *Bedeutungsträger* und es nimmt den Status eines Zeichens an. Ein Zeichen ist etwas, was für etwas anderes steht (Eco), d.h. etwas wie ein Ausdruck, ein Ton, eine Bewegung oder ein Diagramm, das auf etwas anderes verweist, etwa eine Situation, ein physikalisches Objekt, eine Aktion, eine Prozedur, einen anderen Ausdruck. Die Signale, die Bestandteil der Zeichen sind, haben nichts mit der Bedeutung, die sie tragen, gemeinsam. Die Zuordnung zu Bedeutungen innerhalb der Zeichen erfolgt durch Konvention.

Es gibt mehrere Arten von Zeichen. Eine Art sind die linguistischen Zeichen, die aus der Sprache stammen. Diese ist eine Kommunikationsstruktur, die sowohl eine phonetische Artikulation, nämlich die Töne, die zu Worten kombiniert werden, als auch eine grammatische Artikulation, in der Wörter zu Phrasen zusammen gesetzt werden, besitzt. Linguistische Zeichen treten im Allgemeinen in der Form von Zeichenketten auf, und es wird ihnen ein Signifikator zugeordnet, d.h. ein Konzept, eine Idee, die Repräsentation von etwas oder von einem mentalen Zustand.

Eine andere Art von Zeichen sind die Indizes. Sie basieren nicht auf einer Sprache mit einer doppelten Artikulation, sondern beziehen sich auf die Konstruktion von Hypothesen über eine Situation oder eine Sache zurück. Beispiele sind die Spuren von Tieren, die andeuten, dass diese vorbei gegangen sind, die Symptome einer Krankheit, die auf die Krankheit hindeuten, oder der Rauch als Anzeichen für Feuer. Der kognitive Prozess der Ableitung einer Hypothese aus einem Index heißt *Abduktion*. Er ist typisch für untersuchendes Schlussfolgern (z.B. in der Kriminalistik) und für die Diagnose.

6.1.2. Definition und Modelle der Kommunikation

Existierende Ansätze zur Definition von Kommunikation gehen auf die Kommunikationstheorie von Shannon und Weaver (1940) zurück. In dieser Theorie besteht ein Kommunikationsakt aus dem Senden einer Information von einem *Sender* zu einem *Empfänger*. Die Information wird *kodiert* mittels einer *Sprache* und vom Empfänger dekodiert. Die Information wird über einen *Kanal* (oder *Medium*) gesendet. Die Situation, in der sich Sender und Empfänger befinden und sich der ganze Prozess abspielt, heißt *Kontext*. Abbildung 6.1 illustriert diese Vorstellung.

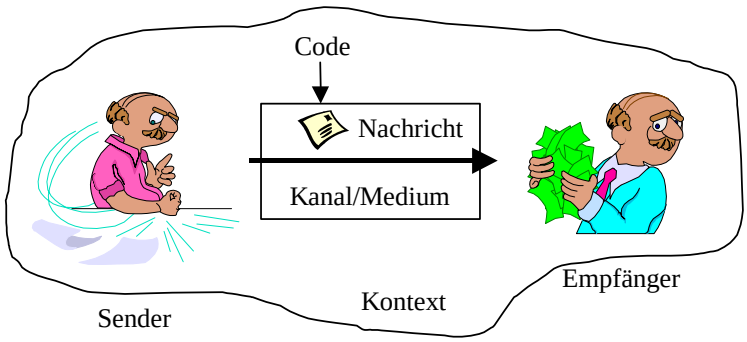

Abbildung 6.1

In den Sozialwissenschaften bildete sich die so genannte *Palo Alto-Schule* aus. Sie betonte, dass Kommunikation als ein integriertes Ganzes sei, d.h. eine Menge bedeutungsvoller Interaktionen zwischen Individuen. Kommunikation ist demnach ein permanenter sozialer Prozess, der verschiedene Verhaltensformen einschließt: Worte, Gesten, Blicke, Mimikry, interpersonaler Raum usw.

Nachrichten haben danach für sich genommen keine Bedeutung, nur der Kontext erhellt die Bedeutung der Interaktionsformen.

Sperber und Wilson (1986) haben ein kognitionswissenschaftliches Modell der Kommunikation entwickelt, in dem die Konzepte der Inferenz und Repräsentation die wesentliche Rolle spielen. Die an der Kommunikation Beteiligten stellen eine Menge von Hypothesen über die Repräsentationen und Ziele des jeweils anderen auf. Der Sender macht eine „minimale“ Äußerung, d.h. eine die er für notwendig hält, damit der Empfänger seine kommunikative Intention versteht. Dabei geht er davon aus, dass der Empfänger auf Grund seiner Repräsentationen aus der Äußerung die richtigen Konsequenzen ableitet.

6.1.3. Kategorien der Kommunikation

In der Linguistik werden Sprechakte dadurch beschrieben, dass Kommunikationsarten mit Bezug auf die Beziehungen zwischen den verschiedenen an der Kommunikation beteiligten Agenten klassifiziert werden. Die Linguistik bezieht sich dabei auf das klassische Kommunikationsmodell. Diese Beziehungen sind:

(1) Die Sender-Empfänger-Verbindung
(2) Die Art des Mediums
(3) Die Kommunikationsintention

Sender-Empfänger-Verbindung

Ist der Empfänger dem Sender bekannt, dann kann er ihm eine spezifische Nachricht schicken und damit eine individuelle Kommunikation initiieren. Man sagt, diese Art der Kommunikation erfolge im *point-to-point*-Modus. Sie ist die bei kognitiven Agenten am häufigsten benutzte. Sie wird in folgender Weise notiert:

⟨id⟩ ⟨Sender⟩ : ⟨Adressat⟩ << ⟨Äußerung⟩

⟨id⟩ ist der (optionale) Identifier der Nachricht, ⟨Sender⟩ der sendende Agent, ⟨Adressat⟩ der Agent, an den die Nachricht gerichtet ist, und ⟨Äußerung⟩ der Inhalt der Nachricht.

Kennt der Sender den Empfänger nicht, dann wird die Nachricht im *broadcasting*-Modus an eine ganze Menge von Agenten verschickt, die zum Empfänger in einer Nachbarschaftsrelation stehen, sei es durch räumliche Nähe, Verknüpfung über ein Netz oder ähnlich. Diese Art der Kommunikation wird vor allem in dynamischen Umgebungen genutzt, in denen Agenten neu entstehen oder verschwinden können. Es wird zum Beispiel in Protokollen zur Aufgabenzuordnung wie dem Contract Net Protcol benutzt. In der obigen Notation wird nur der Empfänger durch das Wort `All` oder durch die Angabe der Nachbarschaftsrelation zwischen Sender und Menge von Adressaten ersetzt.

Broadcasting-Nachrichten können durch eine Menge von point-to-point-Nachrichten ersetzt werden, wenn eine Art vermittelnde Einheit existiert, die das Broadcasting durchführt. So kann z.B. die folgende Broadcasting-Nachricht (M3) durch die Menge von Nachrichten (M3‘) und (M3‘‘) ersetzt werden, wobei `C` die vermittelnde Einheit ist:

```
(M3)    A : {x|P(x)} << M

(M3‘)   A : C << broadcast(M)
(M3‘‘)  for all x of Receivers(C), C : x << M
```

Art des Mediums

Man kann zwischen drei Arten der Nachrichtenübertragung unterscheiden: direkte Übertragung, Übertragung durch Signalpropagierung und Übertragung durch öffentliche Mitteilung.

1. *Direkte Übertragung* erfolgt, wenn ein Agent eine Nachricht dem Kanal übergibt und dieser sie direkt zum Empfänger oder zu den Empfängern transportiert. Die Entfernung wird nicht berücksichtigt, außer dies wäre in der Nachricht spezifiziert, und die Nachricht geht an keine anderen Agenten.

2. *Übertragung durch Signalpropagierung* ist typisch für reaktive Agenten. Ein Agent sendet ein Signal, das in der Umgebung verbreitet wird und dessen Intensität mit zunehmender Entfernung abnimmt. Üblicherweise erfolgt die Abnahme der Intensität linear oder quadratisch mit der Entfernung. Ist x_0 der Ausgangspunkt des Signals, dann hat die Intensität des Signals am Punkt x also einen der beiden folgenden Werte

$$V(x) = \frac{V(x_0)}{dist(x, x_0)} \qquad \text{oder} \qquad V(x) = \frac{V(x_0)}{dist(x, x_0)^2}$$

Eine direkte Kommunikation ist bei dieser Übertragungsart schwierig, denn im Prinzip kann jeder Agent, der nahe genug an der Signalquelle ist, es hören und interpretieren. Soll eine solche Nachricht an einen spezifischen Empfänger gerichtet werden, dann müssen zusätzliche Daten hinzugefügt werden.

Der Effekt, dass die Intensität des ausgesandten Signals mit der Entfernung abnimmt, hat die Wirkung, dass sich räumliche Abstände als soziale Differenzen auswirken, vorausgesetzt die Agenten sind für die Intensität sensitiv. Angenommen, in einem reaktiven Multiagentensystem gebe es eine Quelle, die einen Stimulus S am Punkt x_0 aussendet, und zwei Agenten A und B, in denen S das Verhalten P hervorruft. Wenn sich A näher an der Quelle befindet als B, dann ist das Stimulusniveau für A höher als für B, also wird bei A eine stärkere Tendenz zum Verhalten P aufgebaut als bei B. Gibt es außerdem einen Mechanismus zur Verstärkung bereits ausgelöster Verhaltensweisen, dann der räumliche Unterschied sogar eine unterschiedliche Spezialisierung in den Verhaltensweisen der beiden Agenten hervor rufen. Auf diese Weise führen räumliche Unterschiede zu sozialen Differenzen.

3. *Übertragung durch öffentliche Mitteilung* ist die typische Übertragungsart bei Blackboard-Systemen, wird aber bei Multiagentensystemen seltener benutzt. Ein Agent bringt eine Nachricht, die er kommunizieren will, in einem öffentlichen Raum an, der Nachrichtenbrett, Tafel (Blackboard) oder Äther genannt wird. Diese Übertragungsart kombiniert die direkte Übertragung (durch Angabe einer Adresse im Kopf der Nachricht) mit der Übertragung durch Signalpropagierung.

Kommunikationsintention

Bezüglich der Intention eines Agenten bei einer Kommunikation kann man zwei Formen unterscheiden: *Intentionale Kommunikation* entsteht durch den ausdrücklichen Wunsch eines Agenten zur Kommunikation, der Agent führt dann eine bewusste Handlung aus; *zufällige Kommunikation* entsteht als Nebeneffekt einer anderen Aktion ohne die Absicht des Agenten zur Kommunikation.

Zufällige Kommunikationen haben zwei besondere Merkmale: Ihre Bedeutung ist nur mit dem Zustand des Senders verbunden, auch wenn dieser nicht unbedingt in der Lage ist, die Kommunikation zu steuern, und ihre Interpretation folgt nicht aus einem vorab festgelegten Code, sie ist

vielmehr vollständig variabel und hängt nur vom Empfänger und der Bedeutung, die dieser ihr zuweist, ab. Deshalb können Nachrichten der zufälligen Kommunikation z.B. von einem Tier verwendet werden um Artgenossen zu finden oder von einem Jäger um ein Beutetier zu jagen. Sie haben zwar eine Semantik (sie bezeichnet den Zustand des Senders), aber keine intrinsische Pragmatik, d.h. sie können beim Empfänger unterschiedliches Verhalten auslösen. Darin unterscheiden sie sich grundlegend von der intentionalen Kommunikation, deren Pragmatik durch die Sprechakt-Theorie untersucht werden kann.

Die Intention zur Kommunikation ist nicht ein absoluter Wert, sondern existiert in abgestufter Form, abhängig von den kognitiven Fähigkeiten des Senders. Es gibt komplexe Beziehungen zwischen den Repräsentationsfähigkeiten von Agenten und den Kommunikationsformen, die sie benutzen, wie Tieruntersuchungen ergeben haben. Nach Vauclair gibt es eine Vielzahl von Stufen der Intentionalität, die mit Komplexität der Kommunikation korrespondieren. In dieses System von Intentionalitätsstufen ist die zufällige Kommunikation eingeschlossen.

Intentionalität der Ordnung 0 beschreibt eine Situation, in der der Sender ein Signal sendet, weil er einen bestimmten inneren Zustand aufweist (z.B. wenn er schreit, weil er hungrig ist) oder weil er einen bestimmten Stimulus erhält (z.B. Wahrnehmung eines Jägers), selbst wenn kein Empfänger für das Signal existiert. Die Erzeugung des Signals geht nur auf eine Stimulus/Response-Beziehung zurück und hängt nicht von einer Überlegung seitens des Senders ab. Dies kann in folgender Weise notiert werden

X sendet die Nachricht M weil X S wahrnimmt

wobei S das Signal ist. M ist dann eine *zufällige Nachricht*. Intentionalität der Ordnung 1 setzt einen direkten Wunsch beim Sender voraus eine Wirkung beim Empfänger zu erzielen. Die Relation zwischen Nachricht und Intention kann so formuliert werden:

X sendet die Nachricht M to Y, weil X von Y will, dass er P tut

Höhere Ordnungen von Intention behandeln die Annahmen, die durch Nachrichten erzeugt werden. Zum Beispiel bei Ordnung 2 erwartet der Sender keine Antwort vom Empfänger, sondern will bei ihm eine Annahme über den Zustand der Welt erzeugen. Dies kann so notiert werden:

X sendet die Nachricht M to Y, weil X von Y will, dass er P annimmt

Intention der Ordnung 3 schließt eine dritte Partei ein. Sie wird beschrieben durch

X sendet die Nachricht M to Y, weil X von Y will, dass er annimmt, dass Z P glaubt

In Multiagentensystemen kommen nicht alle möglichen Formen von Kommunikation vor. Tabelle 6.1 zeigt die wichtigsten Formen, dargestellt nach den hier verwendeten Kriterien.

Nachrichtentyp	Kommunikationsmodus	Übertragungsart	Intentionalität
point-to-point symbolische Nachricht	point-to-point	direkt	allgemein intentional
Broadcast symbolische Nachricht	allgemeines Broadcasting	direkt	allgemein intentional
Ankündigung	point-to-point, allgemeines Broadcasting	Nachrichtenbrett	allgemein intentional
Signal	Broadcasting	Propagierung	zufällig

Tabelle 6.1

6.1.4. Zweck der Kommunikation

Man kann Kommunikation auch nach ihrer Funktion klassifizieren, d.h. nach dem, was sie bezwecken soll. Jakobson unterscheidet sechs verschiedene Funktionen: die expressive, conative, referentielle, phatische, poetische und metalinguistische Funktion.

- Die *expressive Funktion* charakterisiert die Haltung des Senders. Sie beschreibt den Zustand des Senders, seine Intentionen und Ziele und wie er die Dinge beurteilt und sieht. Diese Funktion dient dazu, Agenten zu synchronisieren um ihre Aufgaben zu koordinieren und bei ihnen gemeinsame Annahmen zu erreichen.
- Die *conative Funktion* charakterisiert einen Auftrag oder eine Forderung des Senders an den Empfänger. Sie ist mit einer Erwiderung der Akzeptanz oder Ablehnung verknüpft. In den meisten Multiagentensystemen hat die Kommunikation diese Funktion. Sie ist sicherlich die wichtigste Funktion in solchen Systemen.
- Die *referentielle Funktion* bezieht sich auf den Kontext. Sie garantiert das Senden von Daten, die sich auf die Fakten der Welt beziehen. Eine Nachricht mit dieser Funktion bezieht sich deshalb auf den Zustand der Welt oder einer dritten Partei.
- Die *phatische Funktion* dient im Wesentlichen dazu eine Kommunikation zu etablieren, fortzusetzen oder zu unterbrechen. Sie dient insbesondere dazu, das Funktionieren des Kanals zu überprüfen. Eine typische Nachricht mit dieser Funktion ist die Acknowledge-Nachricht in den Protokollen verteilter Systeme.
- Die *poetische Funktion* oder ästhetische Funktion beschreibt eine besondere Betonung der Nachricht. Sie spielt in Multiagentensystemen keine Rolle.
- Die *metalinguistische Funktion* bezieht sich auf alles, was Nachrichten, Sprache und die Kommunikationssituation betrifft. Diese Funktion ist in Multiagentensystemen sehr wichtig, denn sie ermöglicht es, in Nachrichten über Nachrichten zu sprechen und die Konzepte, das Vokabular und die Syntax von Sender und Empfänger zu harmonisieren.

Eine Nachricht besitzt gleichzeitig mehrere Funktionen, insbesondere hat jede Nachricht eine expressive und eine phatische Funktion. Die folgende Nachricht

```
(SM1)     B : A << Für welche x gilt P(x)?
```

hat zunächst conative Funktion (B will A zu einer Antwort veranlassen), sie hat aber auch expressive Funktion (der Zustand des Senders B ist, dass er etwas wissen will), und man kann ihr phatische Funktion beimessen (B wünscht den Dialog fortzusetzen). Die Funktion, die unmittelbar mit einer Nachricht verbunden ist, wird *primäre Funktion* genannt, alle anderen mit ihr verbundenen Funktionen *sekundäre Funktionen*.

6.2. Sprechakte

6.2.1. Sprechen als Handlung

Sprechakte bezeichnen all intentionalen Aktionen, die im Lauf einer Kommunikation ausgeführt werden. Es gibt mehrere Typen von Sprechakten; nach Searle und Vanderveken werden die folgenden Haupttypen unterschieden:

(1) *Assertive* Akte geben Informationen über die Welt, indem sie etwas feststellen.
(2) *Direktive* Akte erteilen dem Adressaten Anweisungen.

(3) *Promissive* Akte verpflichten den Sprecher bestimmte Handlungen in der Zukunft auszuführen.
(4) *Expressive* Akte geben dem Adressaten Hinweise auf den mentalen Zustand des Sprechers.
(5) *Deklarative* Akte sind die bloße Aktion des Machens einer Äußerung.

Für die Verwendung in Multiagentensystemen sind diese Definitionen aber noch zu allgemein. Für sie ist eine präzisere Einteilung erforderlich. Es werden vor allem zwei Veränderungen vorgenommen. Erstens: Die direktiven Akte werden in die interrogativen und exerzitiven Akte unterteilt. Die Akte des ersten Typs stellen Fragen um eine Information zu bekommen, die des zweiten Typs fordern einen Agenten zu einer Aktion in der Welt auf. Zweitens: Es werden Sprechunterakte definiert, und zwar dadurch, dass das Objekt eines Aktes spezifiziert wird, und vor allem durch Charakterisierung der Konversationsstruktur, in die ein Sprechakt eingebettet ist.

6.2.2. Lokutorische, illokutorische und perlokutorische Akte

Sprechakte werden als komplexe Strukturen definiert, die aus drei Komponenten aufgebaut sind:

- Die *lokutorische* Komponente betrifft die materielle Erzeugung von Äußerungen durch das Aussenden von Schallwellen oder das Niederschreiben von Zeichen, also den Modus der Erzeugung von Phrasen mit Hilfe einer Grammatik und eines Lexikons.
- Die *illokutorische* Komponente betrifft die Ausführung des Aktes durch den Sprecher mit bezug auf den Adressaten. Illokutorische Akte lassen sich durch eine illokutorische Absicht (z.B. zusichern, fragen, bitten, versprechen, befehlen, informieren) und einen propositionalen Inhalt, das Objekt dieser Absicht, charakterisieren. Illokutorische Akte kann man in der Form `F(P)` repräsentieren, wobei `F` die illokutorische Absicht und `P` der propositionale Inhalt ist.
- Die *perlokutorische* Komponente betrifft die Wirkungen, die ein Sprechakt auf den Zustand, die Aktionen, Annahmen und Urteile des Adressaten haben kann. Beispiele sind überzeugen, inspirieren, fürchten machen, überreden usw. Für sie gibt es keine Performative, vielmehr sind sie Konsequenzen illokutorischer Akte.

Alle diese Sprechaktkomponenten sind in der selben Äußerung auf verschiedenen Ebenen vorhanden. Deshalb kann man auch von lokutorischen, illokutorischen und perlokutorischen Aspekten einer Äußerung reden. Wenn man Sprechakte diskutiert, dann sollte man sich implizit auf alle drei Aspekte beziehen, nicht nur auf den illokutorischen, wie es z.B. Searle und andere getan haben.

6.2.3. Erfolg und Erfüllung

Erfolg und Erfüllung sind Kriterien für Sprechakte. Zunächst kann ein Sprechakt erfolgreich sein oder nicht. Er ist erfolgreich, wenn sein Zweck erreicht wird. Ein Sprechakt kann auf verschiedene Weise Misserfolg haben:

- Bei der Artikulation des Sprechakts: Unverständliche Aussprache, Rauschen im Übertragungskanal, unbekannte Sprache.
- Bei der Interpretation des Sprechakts: Die Übertragung hat zwar geklappt und den richtigen Adressaten erreicht, aber der interpretiert die illokutorische Absicht des Senders nicht richtig.
- Bei der tatsächlichen Ausführung der gewünschten Handlung: Dafür gibt es eine Fülle von Gründen, z.B. das Fehlen der erforderlichen Fähigkeit oder die Weigerung des Adressaten. Der Adressat muss auch bereit sein, sich auf Versprechungen des Senders zu verlassen.

Allgemein ist ein Sprechakt erfolgreich, wenn der Sender den illokutorischen Akt ausführt, der mit dem Sprechakt verbunden ist. Zum Beispiel ist ein Versprechen erfolgreich, wenn der Sender sich verpflichtet, das Versprochene zu tun, oder eine Erklärung ist erfolgreich, wenn der Sender die Autorität hat das Erklärte auszuführen.

Im Unterschied zum Kriterium des Erfolgs bezieht sich das Kriterium der Erfüllung auf die perlokutorische Komponente des Sprechakts und berücksichtigt den Zustand der Welt, der aus dem Sprechakt folgt. So ist eine Bitte etwas zu tun erfüllt, wenn der Adressat ihr nachkommt, eine Frage ist erfüllt, wenn der Adressat darauf antwortet, eine Zusicherung ist erfüllt, wenn sie wahr ist, ein Versprechen ist erfüllt, wenn es gehalten wird. Das Kriterium der Erfüllung ist stärker als der Erfolg, denn aus der Erfüllung eines Sprechakts folgt der Erfolg, aber nicht umgekehrt.

6.3. Konversationen

Ein Sprechakt kommt nicht isoliert vor, sondern er hat in der Regel andere (Sprech-) Akte zur Folge. Ein Versprechen etwas zu tun erwartet die Ausführung der betreffenden Handlung, eine Bitte erwartet eine Handlung und hat weitere Aktionen zur Folge, z.B. Zustimmung oder Verweigerung durch den Adressaten und eine Benachrichtigung des Senders. Solche nachfolgenden Sprechakte sind für den Sender von Bedeutung, denn sie beeinflussen sein weiteres Handeln und seine Erwartungen.

Ein Sprechakt löst in der Regel weitere Sprechakte aus, führt also zu einer Kommunikation. Der Sprechakt

```
(M7) A : B << Question(welche Elemente enthält die Menge {x|P(x)})
```

erwartet z.B. eine der folgenden drei Reaktionen:

(1) Die Antwort auf die Frage:

(M8) `B : A << Answer(M7, a1,…,an)`

wobei $a_1,\ldots,a_n$ die Werte sind, die `P(x)` erfüllen.

(2) Eine Weigerung zur Beantwortung der Frage, eventuell zusammen mit einer Erklärung:

(M8') `B : A << RefuseIncompetentRequest(M7)`

(3) Eine Bitte um weitere Information, wodurch der Dialog auf eine Metaebene gehoben wird, denn jetzt muss über die Kommunikation gesprochen werden:

(M8") `B : A << MetaQuestion(M7, Argumente(P))`

Um solche Konversationen zu modellieren benötigt man *Protokolle*, d.h. gültige Folgen von Sprechakten oder Nachrichten. Als Darstellungsmittel für Protokolle können z.B. endliche Automaten oder Petrinetze verwendet werden.

6.3.1. Konversationen und endliche Automaten

Konversationen können als gerichtete Graphen beschrieben werden. Die Knoten repräsentieren Zustände der Kommunikation, die Kanten Übergänge, die durch Sprechakte oder Nachrichten ausgelöst werden. Ein solcher Graph kann auch als Zustandsübergangsgraph, also als endlicher Automat betrachtet werden. Der Graph von Abbildung 6.2 ist ein Beispiel für ein Protokoll.

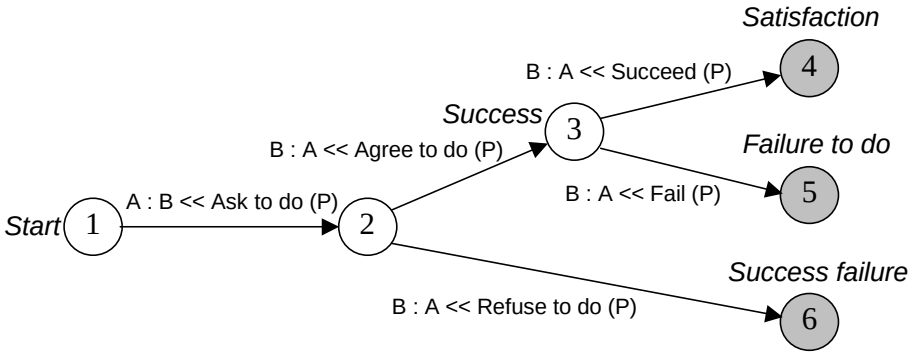

Abbildung 6.2

Die wichtigsten Merkmale einer kommunikativen Struktur sind:

(1) Die Konversation beginnt mit einem größeren Sprechakt (zusichern, versprechend, fragend, erklärend), der als Folge der Intention eines Agenten gegenüber einem anderen Agenten geäußert wird.
(2) In jedem Zustand der Konversation gibt es eine eingeschränkte Menge möglicher Aktionen. Diese sind größere oder kleinere Sprechakte, z.B. Weigerung, Zustimmung usw.
(3) Es gibt Endzustände, die das Ende der Konversation bestimmen. Wird einer dieser Zustände erreicht, dann wird die Konversation als abgeschlossen betrachtet.
(4) Ein Sprechakt verändert nicht nur den Zustand der Konversation, sondern auch den Zustand der Agenten (Annahmen, Verpflichtungen, Intentionen), der durch die Konversation hervorgerufen wird.

6.3.2. Konversationen und Petrinetze

Verwendet man Petrinetze zur Modellierung von Protokollen in Multiagentensystemen, dann wird jeder Agent durch ein eigenes Petrinetz beschrieben. In diesem werden seine internen Zustände durch Plätze und die durch die Konversation hervorgerufenen Übergänge zwischen ihnen durch Transitionen dargestellt. Genauer dienen die Transitionen teilweise zur Synchronisation der eingehenden Nachrichten, teilweise repräsentieren sie die Bedingungen von Aktionen. Die Sprechakte werden durch Plätze zwischen den beiden Agentennetzen repräsentiert. Abbildung 6.3 stellt die Konversation von Abbildung 6.2 in Form eines Petrinetzes dar.

Der Vorteil der Petrinetze ist, dass sie nicht nur die Sprechakte darstellen, sondern gleichzeitig auch die internen Zustände der Agenten mit ihren Übergängen. Dadurch kann genauer beschrieben werden, was während der Konversation abläuft. Auch die typischen Merkmale der Sprechakte können dargestellt werden, z.B. die Notation von Erfolg, Misserfolg oder Erfüllung. Mit Hilfe gefärbter Petrinetze kann auch die gleichzeitige Konversation mehrerer Agenten modelliert werden. Die Tokens sind dann Strukturen, die den Inhalt der Konversation widerspiegeln. Die einzelnen Sprechakte werden nummeriert, so dass sie jeweils der richtigen Konversation zugeordnet werden können.

6.3.3. Eine Klassifikation der Sprechakte in Multiagenten-Konversationsstrukturen

Die Beschreibung eines Sprechakts besteht aus zwei Teilen: Einem Formular, in dem die einzelnen Angaben zum Sprechakt enthalten sind, und einem BRIC-Diagramm, das die konversationale Struktur wiedergibt. In den Formularen sind die Bedingungen für die Anwendung des Sprechakts aus der Sicht des Initiators enthalten, die Bestandteile der illokutorischen Komponente des Sprechakts, die Bedingungen für Erfolg und Erfüllung, mögliche Fehler und normale Konsequenzen. Das Formular besteht aus folgenden Teilen:

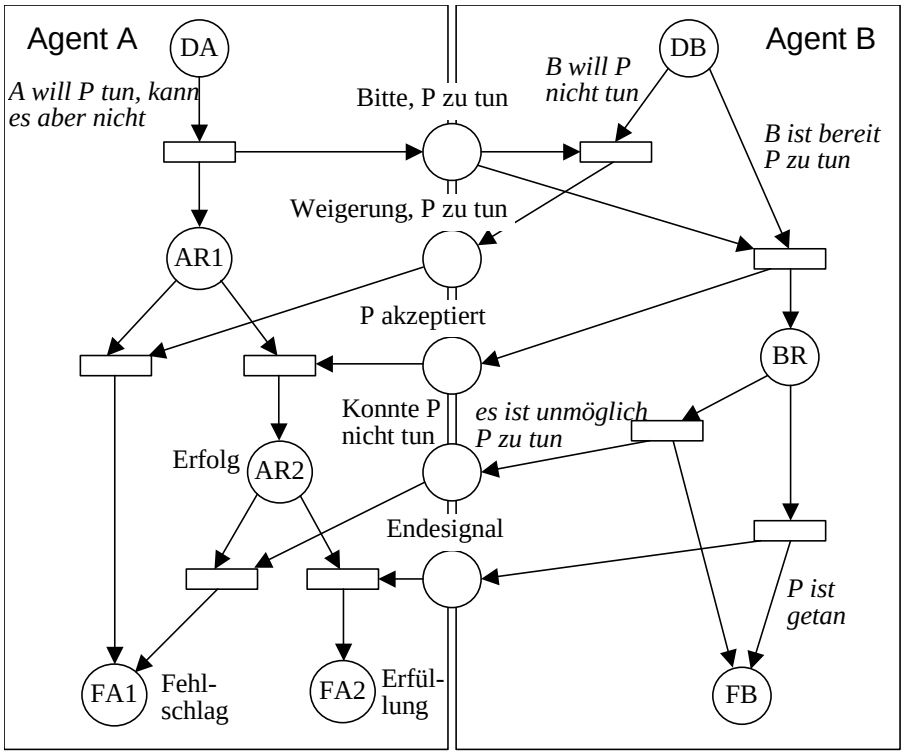

Abbildung 6.3

Format: Beschreibt die Syntax der Nachricht, die mit dem Sprechakt verbunden ist. Es hat die Form `A : B << F(P1,...,Pn)`, wobei `A` der sendende Agent ist, `B` der Adressat, `F` das Performativ, das den Sprechakt anzeigt, und wobei die P_i den propositionalen Inhalt des Sprechakts bezeichnen.

Subakte: Listet die Varianten des Sprechakts und ihre Performative auf.

Bedingungen: Gibt die Bedingungen für die Realisierung des Sprechakts an, d.h. die Bedingungen, die erfüllt sein müssen für die Anwendung. Sie beziehen sich allgemein auf die Intentionen und Annahmen des Senders und auf die Fähigkeiten der Kommunikationspartner.

Erfolg: Beschreibt die normalen Konsequenzen des Erfolgs des Sprechakts und die damit verbundene Nachricht.

Erfüllung: Listet die Nachbedingungen der Erfüllung und die damit verbundene Nachricht auf.

Misserfolg: Beschreibt die erfassbaren Bedingungen eines Misserfolgs, d.h. diejenigen, die auf Grund des Nichterfülltseins einzelner Bedingungen der Realisierung vorhersagbar sind. Misserfolge, die mittels spezifischer Nachrichten mitgeteilt werden, haben Folgen für die Annahmen des Senders und seinen Zustand. Erklärungen werden nicht gegeben, sie müssen bei Bedarf extra abgefragt werden.

Die Bedingungen und Konsequenzen der Sprechakte werden mit Hilfe einiger grundlegender Cognitons und Prädikate formuliert, die im Folgenden aufgelistet sind.

Kognitons

`believe(A,P)`: Agent `A` nimmt `P` an. Annahmen werden als überprüfbar betrachtet. `believe(A,P)` in prädikativer Verwendung bedeutet, dass Agent `A` annimmt, dass `P` gilt, innerhalb einer Aktion, dass Agent `A` ab sofort annimmt, dass `P` gilt und seine Annahmen dementsprechend aktualisiert.

`goal(A,P)`: Agent `A` hat das Ziel, dass Bedingung `P` wahr sein soll.

`intention(A,P)`: Agent `A` hat die Intention, die Aktion `P` auszuführen.

`commit(A,B,P)`: Agent `A` ist verpflichtet `P` für `B` zu tun.

`competent(A,P)`: Agent `A` hat die Fähigkeit Aktion `P` auszuführen.

Der Satz `believe(A,competent(B,P))` besagt, dass `A` glaubt, `B` habe die Fähigkeit `P` auszuführen.

Prädikate

`compute(E)`: Repräsentiert die Aktion einen Ausdruck `E` zu berechnen.

`perceive(A,E)`: Zeigt an, ob Agent `A` das Objekt oder die Situation `E` wahrnimmt.

`exec(P)`: Zeigt an, ob Aktion `P` ausgeführt wird oder nicht.

`wantDo(A,P)`: Agent `A` wird irgendwann in der Zukunft bereit sein, `P` zu tun.

Ausführungs-Sprechakte

Die wichtigsten Ausführungs-Sprechakte sind Delegation und Subskription. Beide dienen zum Initiieren einer Konversation.

Delegation

Delegationen sind z.B. Bitten um Aktionen oder Bitten um Lösungen. Sie werden neben den Fragen am häufigsten zur Eröffnung einer Konversation benutzt. Eine Bitte wird an einen anderen Agenten geschickt und dieser informiert darüber, ob er sich zur Ausführung der Aufgabe verpflichtet oder nicht. Wenn sich der andere Agent verpflichtet, dann kann er auf zwei Arten reagieren:

a) Er sendet eine Nachricht über das Ende der Aktion zurück oder schickt die Lösungen zurück, falls es sich um ein zu lösendes Problem handelt.

b) Er informiert, dass es ihm nicht möglich war seine Verpflichtung einzulösen, trotz seines Versprechens.

Abbildung 6.4 zeigt ein Formular und ein BRIC-Diagramm für die Delegation.

Bitten: Grundlegender Sprechakt, bei dem ein anderer Agent um die Ausführung einer Aufgabe (Ausführen einer Aktion oder Lösen eines Problems) gebeten wird.

Format

`A : B << Request(P)` wobei P eine Aufgabe ist

Subakte

`RequestSolve(P)` wobei P ein Problem ist

Bedingungen

`goal(A,C) ∧ believe(A,exec(P)⇒C) ∧ believe(A,competent(B,P)) ∧ believe(A,wantDo(A,P)) ∧ [believe(A,¬competent(A,P)) ∨ ¬wantDo(A,P)]`

`believe(A,¬eventually(C))`

Erfolg

`A : B << Request(P) → commit(B,A,P)`

Erfüllung

`B : A << NotificationEndAction(P) → believe(A,C)` wobei `P` eine Aktion ist

`B : A << Result(P,R) → believe(A,R)` wobei `P` ein Problem ist

Misserfolg

`B : A << RefuseDo(P) → believe(A,¬competent(B,P)) ∨ ¬wantDo(B,P)`

`A << FailureDo(P) →` frage nach Erklärungen

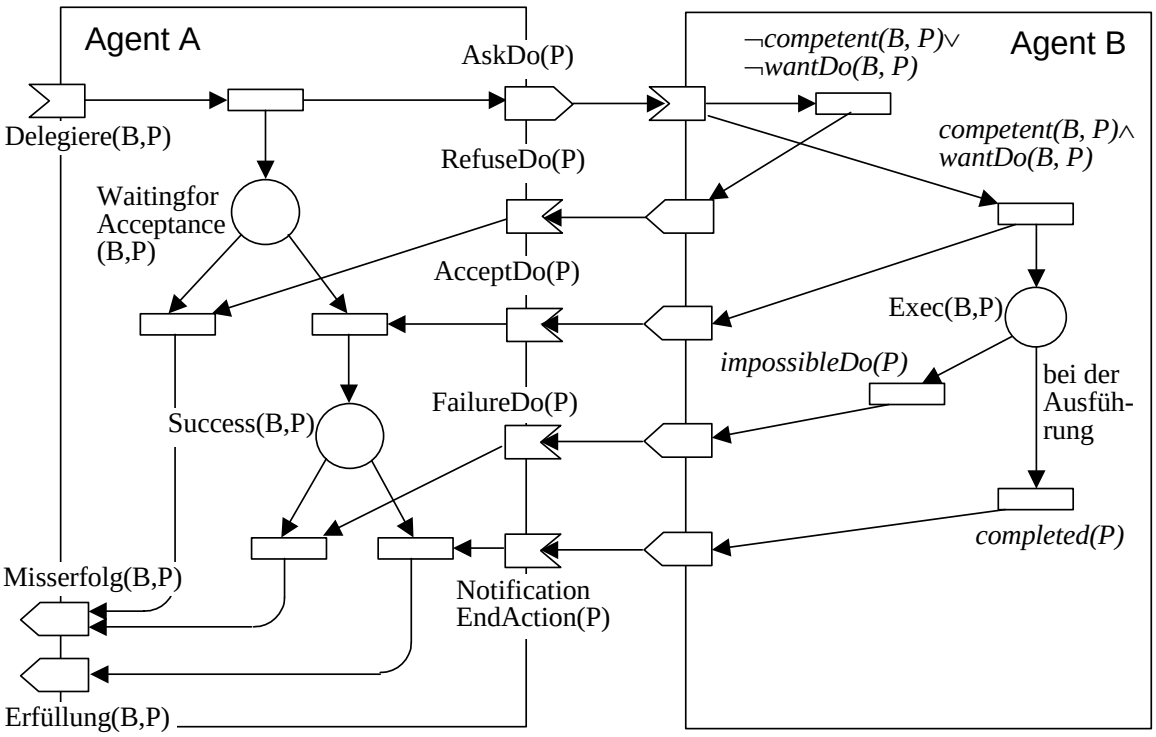

Abbildung 6.4

Subskription

Mit Subskription ist die Bitte eines Agenten um Information über das Eintreten von Ereignissen oder über das Wissen eines anderen Agenten über einen speziellen Gegenstand. Tätigkeiten, die etwas mit Erforschung, Überwachung oder Wahl zu tun haben sind Beispiele dafür, wo Subskription vorkommen kann, wenn nämlich ein Agent die Aufgabe hat, Informationen über die Ereignisse an andere zu schicken. Die Subskription kann in regelmäßigen Abständen oder ereignisorientiert erfolgen. Abbildung 6.5 zeigt ein Formular und ein BRIC-Diagramm für die Subskription.

Abbonnieren: Grundlegender Sprechakt, bei dem ein Agent A einen Agenten B bittet ihn über etwas zu informieren, und zwar entweder in regelmäßigen Abständen, wenn ein Ereignis stattfindet oder wenn B irgendwelche neue Informationen hat.

Format

`A : B << Subscribe(P,R)` wobei `P` eine Bedingung und `R` eine Aufgabe ist

Bedingungen

`goal(A,believe(B,P)) ⇒ believe(A,R))`

`believe(A,eventually(believe(B,P))) ∧ believe(A,believe(B,P)) ⇒`
`wantDo(B,B : A << Advise(R)))`

Erfolg

`B : A << AcceptSubscribe(P,R) →`
`commit(B,A,believe(B,P) ⇒ ¬wantDo(B,B : A << Advise(R)))`

Erfüllung

`B : A << Advise(R) → believe(A,R)`

Misserfolg

`B : A << RefuseSubscribe(P,R) →`
`believe(A,¬wantDo(B,believe(B,P)⇒ wantDo(B,B : A << Advise(R)))`

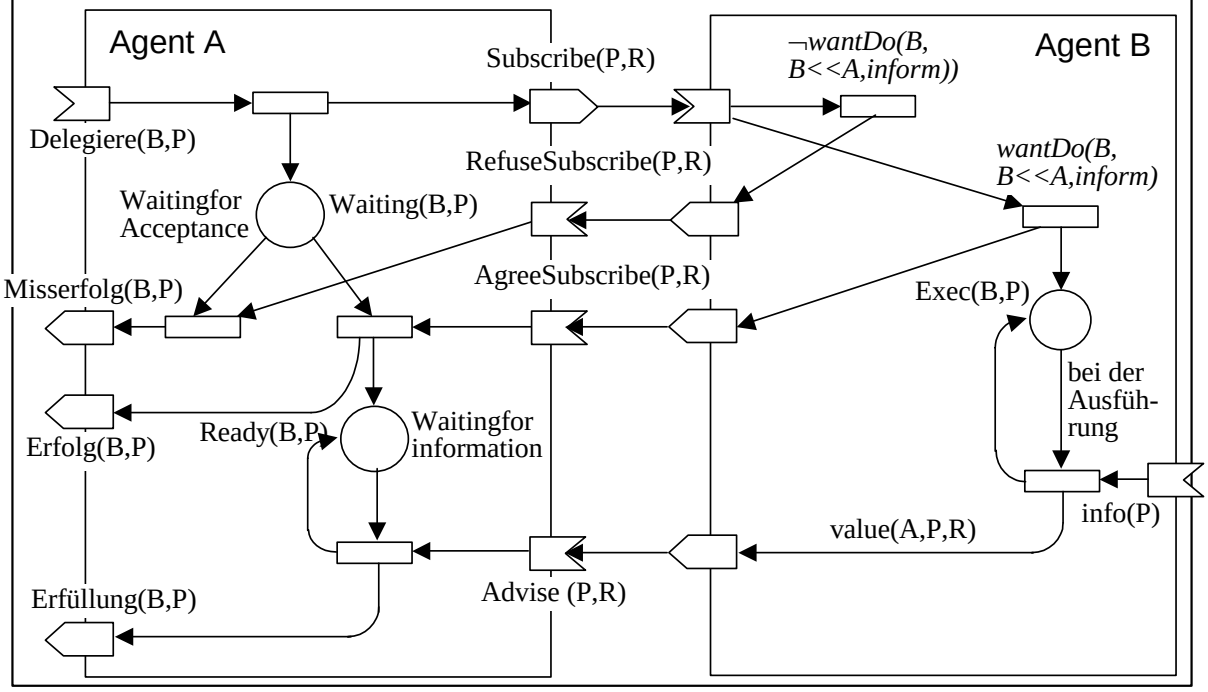

Abbildung 6.5

Interrogative Sprechakte

Interrogative Sprechakte werden für Fragen benutzt, deren Beantwortung keine komplexe Berechnung erfordert. Die Fragen können sich auf die Welt beziehen, z.B. Fragen danach, was der Agent annimmt, oder auf den mentalen Zustand des Adressaten, z.B. nach seinen Zielen, Verpflichtungen, Kontakten usw. Abbildung 6.6 zeigt ein Formular und ein BRIC-Diagramm für interrogative Sprechakte.

Fragen: Grundlegender Sprechakt, bei dem ein anderer Agent um Information gebeten wird.

Format

`A : B << Question(f,P)` wobei `P` ein Datum und `f` eine Funktion ist

Subakte

`QuestionBoolean(P)` wobei `P` eine Aussage ist

`QuestionReasons(E)` wobei `E` ein Ausdruck ohne Variable ist

Bedingungen

`goal(A,believe(A,R)` mit

`R = f(P) ∧ believe(A,competent(B,f) ∧ believe(A,¬competent(A,f))`
`∧ believe(A,¬believe(A,R)) ∧ believe(A,wantAnswer(B,f))`

Erfüllung

`B : A << Answer(R)` → `believe(A,R)`

Misserfolg

`B : A << RefuseAnswer(R)` →
`believe(A,¬wantAnswer(f) ∨ ¬competent(B,f))`

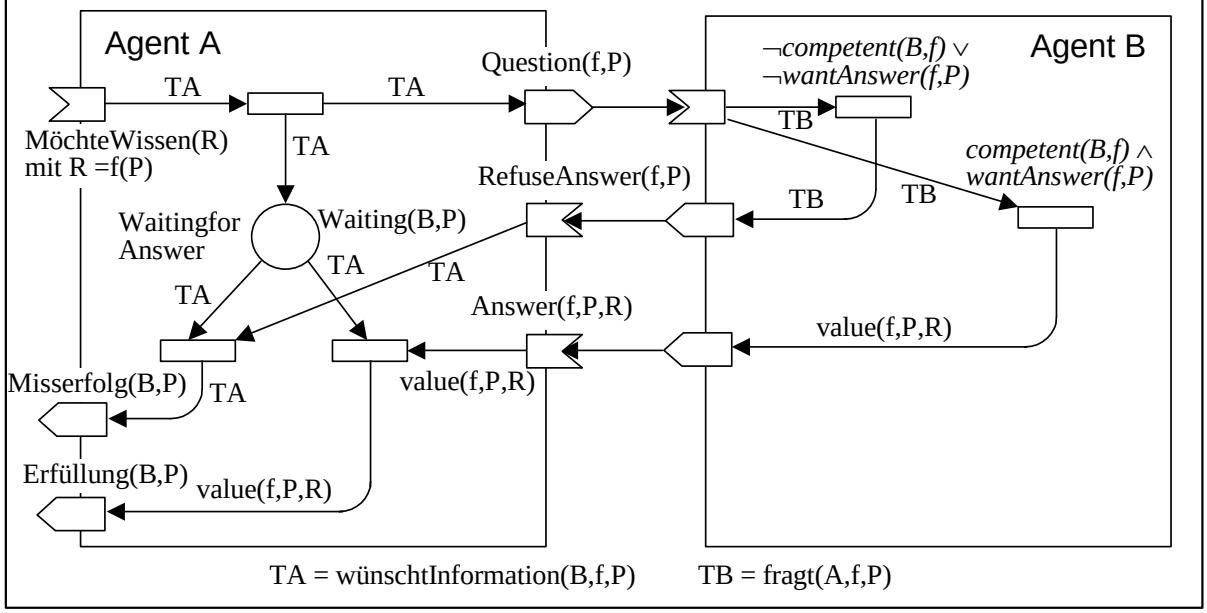

Abbildung 6.6

Assertive Sprechakte

Assertive Sprechakte dienen dazu, Informationen über die Welt, also über Objekte oder Situationen, zu übermitteln. Man kann zwei Ebenen assertiver Sprechakte unterscheiden: Bei der ersten werden Informationen als Antwort auf Fragen übermittelt; sie sind deshalb Bestandteil anderer Sprechakte, etwa der interrogativen. Bei der zweiten möchte man einen anderen Agenten von etwas überzeugen. Im ersten Fall sind Erfolg und Erfüllung miteinander verbunden, vgl. die Formalisierung des interrogativen Sprechakts. Im zweiten Fall besteht der Sprechakt aus zwei Phasen: In der ersten werden die Daten gesendet, sie kann zum erfolg führen, in der zweiten werden die Daten akzeptiert, sie kann zur Erfüllung führen, und zwar dann, wenn die Information mit den Annahmen des Empfängers kompatibel ist. Abbildung 6.7 zeigt ein Formular und ein BRIC-Diagramm für den Sprechakt *Zusichern*.

Zusichern: Grundlegender Sprechakt, bei dem ein anderer Agent von etwas überzeugt werden soll.

Format
`A : B << Assert(R)`

Bedingungen
`goal(A,believe(B,R)) ∧ believe(A,¬believe(B,R))`

Erfüllung
`B : A, Acquiesce(R) → believe(A,believe(B,R))`

Misserfolg
`B : A, RefuseBelieve(R) → believe(A,believe(B,¬R))`

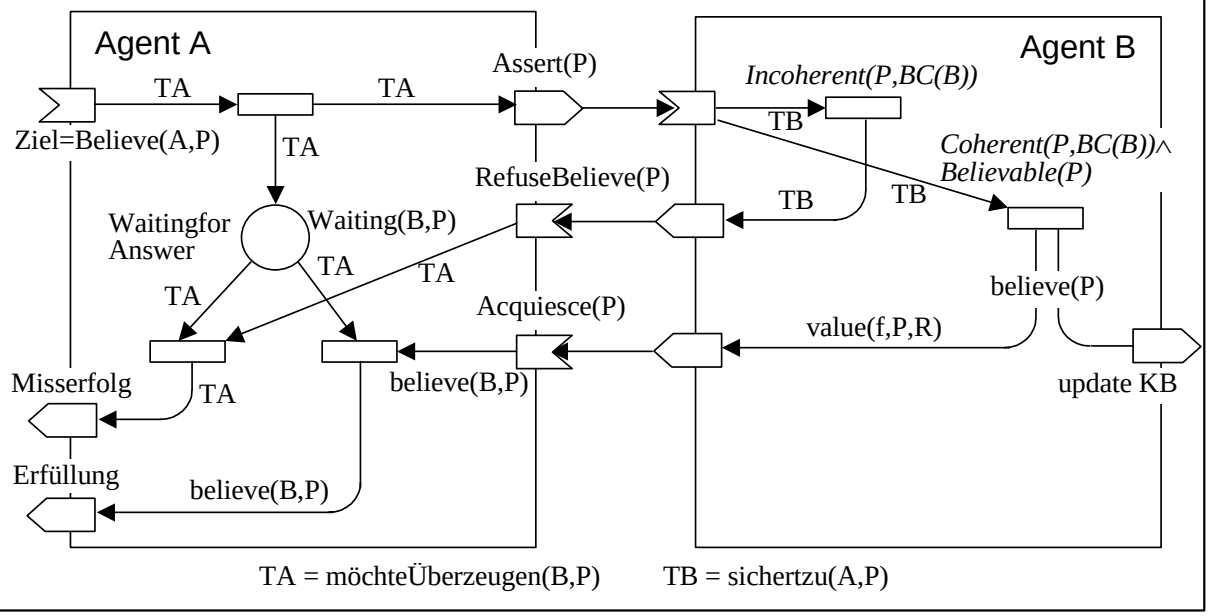

Abbildung 6.7

Zusätzlich zu den Hauptsprechakten, die bisher behandelt wurden, kann man noch weitere Sprechakte definieren, u.a. die folgenden:

Promissive Sprechakte

`OfferService`: Ein Agent `A` bietet einem anderen Agenten `B` seine Dienste an. `B` kann sie ablehnen oder akzeptieren. `A` verpflichtet sich dabei gegenüber `B` und muss, seinen Fähigkeiten entsprechend, die Befehle von `B` ausführen.

`Propose/Confirm/Refute/Hypothesis`: Diese Performative werden beim verteilten Problemlösen benutzt.

Expressive Sprechakte

`KnowHowDo`: Ein Agent `A` zeigt `B` seine Fähigkeiten an.

`Believe`: Ein Agent `A` zeigt `B` seine Überzeugung über einen Sachverhalt an.